Topological Analysis of Biological Pathways:

Genes, MicroRNAs and Pathways Involved in Hepatocellular Carcinoma

by

Chaoxing Li

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2017 by the
Graduate Supervisory Committee:

Valentin Dinu, Co-chair
Yang Kuang, Co-chair
Li Liu
Xiao Wang

ARIZONA STATE UNIVERSITY

December 2017

ABSTRACT

Rewired biological pathways and/or rewired microRNA (miRNA)-mRNA interactions might also influence the activity of biological pathways. Here, rewired biological pathways is defined as differential (rewiring) effect of genes on the topology of biological pathways between controls and cases. Similarly, rewired miRNA-mRNA interactions are defined as the differential (rewiring) effects of miRNAs on the topology of biological pathways between controls and cases. In the dissertation, it is discussed that how rewired biological pathways (Chapter 1) and/or rewired miRNA-mRNA interactions (Chapter 2) aberrantly influence the activity of biological pathways and their association with disease.

This dissertation proposes two PageRank-based analytical methods, Pathways of Topological Rank Analysis (PoTRA) and miR2Pathway, discussed in Chapter 1 and Chapter 2, respectively. PoTRA focuses on detecting pathways with an altered number of hub genes in corresponding pathways between two phenotypes. The basis for PoTRA is that the loss of connectivity is a common topological trait of cancer networks, as well as the prior knowledge that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs. However, from normal to cancer, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network, namely, the number of hub genes might be altered in cancer compared to that in normal samples. Hence, it is hypothesized that if the number of hub genes is different in a pathway between normal and cancer, this pathway might be involved in cancer. MiR2Pathway focuses on quantifying the differential effects of miRNAs on the activity of a biological pathway when miRNA-mRNA connections are altered from

normal to disease and rank disease risk of rewired miRNA-mediated biological pathways. This dissertation explores how rewired gene-gene interactions and rewired miRNA-mRNA interactions lead to aberrant activity of biological pathways, and rank pathways for their disease risk. The two methods proposed here can be used to complement existing genomics analysis methods to facilitate the study of biological mechanisms behind disease at the systems-level.

DEDICATION

To my father (Yongfu Li), mother (Xiaohua Rao) and brother (Chaozheng Li) who always support me without whom this dissertation might not be completed sooner.

ACKNOWLEDGMENTS

I am profoundly grateful to co-chair Drs. Valentin Dinu and Yang Kuang without whom none of my projects would have been possible. Also, I am very thankful to other committee members: Drs. Li Liu and Xiao Wang. Specifically, Dr. Li Liu guided me in one of my projects -- Pathways of Topological Rank Analysis (PoTRA): A Novel Method to Detect Pathways Involved in Cancer. My Ph.D committee members not only provided encouragement and continuous guidance, but also challenged me for critical thinking and self-motivated research. Their advice and encouragement really motivated me to explore and overcome difficulties in my research activities. At the same time, they gave me much space to exert my independence on thinking and conducting research. Most importantly, I thank Dr. Valentin Dinu for prompt feedbacks and timely suggestions to my research questions and approaches, drafts of our manuscripts and this thesis. Also, I am greatly indebted to Dr. Liu for giving the timely guidance and great suggestions when I came across the difficulties on one of my research projects.

I would also like to thank Dr. Carlo Maley for providing me a research assistant position in Biodesign Institute. I learned a lot from him and the many colleagues in his lab on critical thinking and how to overcome difficulties on research.

Last but not least, I am forever grateful to my parents and brother for their constant love, unconditional spiritual support and confidence in me.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

Chapter 1: INTRODUCTION

1.1. Alteration in expression is associated with diseases

It is well known that genomic alterations might lead to diseases. There have been many studies demonstrating it. For example, Hemophilia A and Huntington disease (HD) are monogenic diseases, which suggests that they might be caused by alteration of one gene. In addition to monogenic diseases, many other diseases might be caused by a combination of genomic alterations, epigenetic, miRNA-mediated, environmental and lifestyle factors, which are called complex diseases, such as diabetes, schizophrenia, or cancer.

Several methods have been developed to identify active subnetworks from expression changes between two phenotypes. For example, Gene Set Enrichment Analysis (GSEA) is a well-known analytical method that determines whether a pre-defined set of genes shows statistically significant, concordant differences between two different phenotypes, which is also based on differential expression of a set of genes between two different phenotypes (Subramanian et al., 2005).

Besides, there is another factor, miRNA, able to cause aberrant activity of genes, thereby be associated with diseases. MiRNAs are short non-coding RNAs of about 22 nucleotides in length, involved in the post-transcriptional regulation of gene expression. MiRNAs can induce the degradation of mRNA or translational repression of mRNA depending on the degree of homology to specific sequences, typically in the untranslated regions (UTRs) of their targets (Pasquinelli, 2012). MiRNAs can influence the expression of one or many genes at a time. It is reported that more than 60% of human genes are regulated by miRNAs (Friedman et al., 2009). Thus, miRNAs are regarded as key regulators of biological processes. That is to say, a single miRNA can control a biological process by

simultaneously targeting multiple genes of this specific biological process, and those targeted genes might be members of a cascade functioning towards a functional endpoint in the same biological pathways or in the crosstalk between biological processes (Lima et al., 2011). Over recent years, there are many studies demonstrating the role of miRNAs on controlling a wide variety of fundamental biological processes involved in proliferation, apoptosis, cell growth, differentiation, invasiveness, motility, and other oncogenic processes. For example, miRNA-21 can down-regulate the activity of the IL-12/IFN-γ pathway in lung cancer (Lu et al., 2011). MiRNA-7 can influence activity of the PI3-kinase/Akt pathway in hepatocellular carcinoma and glioblastoma (Kefas et al., 2008; Fang et al., 2012). MiRNA-200 is able to influence activity of E-cadherin and Wnt/β-catenin signaling pathways (Saydam et al., 2009). Based on those, several tools have been developed to detect miRNA-pathway associations (Nam et al., 2009; Maragkakis et al., 2011; Hsu et al., 2011a; Lu et al., 2012; Ben-Hamo & Efroni, 2015; Godard & van Eyll, 2015; Preusse, Theis & Mueller, 2016; Backes et al., 2016; Han et al., 2016; Backes et al., 2017).

Indeed, differential gene expression level in biological network might influence phenotypes. However, only investigating the differential expression levels of gene may be not sufficient since the topology of biological pathway is also an important characteristic of biological pathways and the role they play in both normal and pathological processes, as described below.

1.2. Alteration in topology is associated with diseases

Each biological pathway exerts its function by delivering signaling through the gene network. Theoretically, a pathway is supposed to have a robust topological structure

under normal physiological conditions. However, the pathway's topological structure could be altered under some pathological condition. It is well known that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs. In addition, it is reported that the loss of connectivity is a common topological trait of cancer networks. Hence, from normal to cancer, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network, namely, the number of hub genes might be altered in cancer compared to that in normal. Based on this, we propose a new PageRank-based method called Pathways of Topological Rank Analysis (PoTRA) to detect pathways involved in cancer.

Recent studies have shown miRNAs as key regulators of a wide variety of biological processes, such as proliferation, differentiation, apoptosis, metabolism, etc. Rewired miRNA-mRNA connections can influence the activity of biological pathways, because miRNA-mRNA connections tend to be dynamic or condition-specific, or differential between disease and non-disease. Here, we define rewired miRNA-mRNA connections as the differential (rewiring) effects on the activity of biological pathways between diseased and normal phenotypes. Chapter 3 proposes a PageRank-based method to measure the degree of miRNA-mediated dysregulation of biological pathways between HCC and normal samples based on rewired miRNA-mRNA connections. The degree of miRNA-mediated dysregulation of biological pathways is regarded as disease risk of biological pathways by measuring total differential influence of all miRNAs on the activity of a pathways between diseased and normal conditions, thereby I can rank biological pathways for disease risk.

1.3. Google search algorithm: PageRank

Chapter 2 and Chapter 3 propose two methods, PoTRA and miR2Pathway, which are based on PageRank. The PageRank algorithm is used by the Google search engine to rank the importance of web pages, which is based on the assumption that the importance of a web page is high in a network if this web page has connections with other nodes of high importance. This idea is naturally applied to analyzing biological networks, where the importance of a gene is high if this gene is connected to other genes of high importance. In our study, the gene-gene network is an undirected graph where a node represents a gene and the edges can be defined by prior knowledge (e.g., KEGG database).

The output from the PageRank algorithm is a probability distribution representing the likelihood that a person randomly clicking on links will arrive at any particular web page. A probability is a numeric value between 0 and 1. The sum of probabilities for all web pages is equal to 1. The probability of a web page is proportional with the time spent at the web page when a person surfs the web. This idea can also be intuitively extended to ranking genes in gene networks where the probability of a gene is proportional with the time a research scientist spends looking and returning at the same gene when analyzing research results. For additional details of PageRank, please refer to (Page et al., 1999).

1.4. Summary

In summary, rewired gene-gene interactions and miRNA-mRNA interactions might cause aberrant activity of biological pathways. Chapter 2 focuses on how rewired gene-gene interactions lead to aberrant activity of biological pathways, and Chapter 3 focuses on how rewired miRNA-mRNA interactions lead to aberrant activity of biological pathways. In Chapter 2, each biological pathway exerts its function by delivering

signaling through the gene network. Theoretically, a pathway is supposed to have a robust topological structure under normal physiological conditions. However, the pathway's topological structure could be altered under some pathological condition. It is well known that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs. In addition, it is reported that the loss of connectivity is a common topological trait of cancer networks. Hence, from normal to cancer, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network, namely, the number of hub genes might be altered in cancer compared to that in normal. Based on this, we propose a new PageRank-based method called Pathways of Topological Rank Analysis (PoTRA) to detect pathways involved in cancer. We use PageRank to measure the relative topological ranks of genes in each biological pathway, then select hub genes for each pathway, and use Fisher's exact test to test if the number of hub genes in each pathway is altered from normal to cancer. We apply PoTRA to study hepatocellular carcinoma (HCC) and several subtypes of HCC. In Chapter 3, I present a PageRank-based method, called miR2Pathway, to measure the degree of miRNA-mediated dysregulation of biological pathways between HCC and normal sample based on rewired miRNA-mRNA connections. miR2Pathway proposed here systematically shows the first evidence for a mechanism of biological pathways dysregulated by rewired miRNA-mRNA connections,

Chapter 2: Pathways of Topological Rank Analysis (PoTRA): A Novel Method to Detect Pathways Involved in Cancer

2.1. Introduction

High throughput technologies, such as genomic sequencing and microarrays, allow the genome-wide analysis of molecular factors associated with disease. While the technologies have advanced and have been refined to generate an increasing amount of high quality data, challenges remain in understanding the biological processes involved in disease etiology, particularly for complex disorders.

As we know, individual genomic alterations may result in diseases. For example, Hemophilia A is an X-linked recessive bleeding disorder caused by a deficiency in the activity of coagulation factor VIII (Franchini & Mannucci, 2012). Huntington disease (HD) is an autosomal dominant progressive neurodegenerative disorder with a distinct phenotype characterized by chorea, dystonia, incoordination, cognitive decline, and behavioral difficulties, which is caused by a heterozygous expanded trinucleotide repeat (CAG)n, encoding glutamine, in the gene encoding huntingtin (HTT) on chromosome 4p16 (Walker, 2007; Dayalu & Albin, 2015).

In addition to monogenic diseases, many diseases are complex, such as diabetes, schizophrenia, or cancer, and are believed to be caused by a combination of genomic alterations, epigenetic, environmental and lifestyle factors (Schork, 1997; Hindorff, Gillanders & Manolio, 2011). Genomic disease association analysis suggests that complex diseases are not caused by individual genomic alterations. First, the complex disease phenotype is associated with many genes. Second, it may be associated with interactions among many genes. Therefore, more and more literature has been focusing on analyzing

sets of genes associated with some phenotype. Gene expression profiles have been used to assess the activity of biological networks. Several approaches have been developed to identify active subnetworks across different phenotypes from changes in gene expression. One of the first such studies is a general approach to searching for "active sub-networks" associated with high levels of differential expression (Ideker et al., 2002). This approach identifies a set of genes that form a subnetwork whose expression is altered across two different phenotypes. Another very well-known method, Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005), is a computational method that determines whether a pre-defined set of genes shows statistically significant, concordant differences between two phenotypes, which is also based on differential expression of a set of genes between two phenotypes. These approaches, while powerful and popular, are limited by the fact that they ignore the topology of the gene networks and sets that they investigate. Indeed, differential gene expression level in biological network might influence phenotypes. However, only investigating the differential expression levels of gene may be not sufficient since the topology of biological pathway is also an important characteristic of biological pathways and the role they play in both normal and pathological processes, as described below.

It is well known that the topological structure is very important for biological networks and it determines how genes interact with each other, governing how specific genes and biological pathways operate in the promotion or inhibition of human diseases (Tavazoie et al., 1999; Goeman & Bühlmann, 2007; Tarca et al., 2009; Taylor et al., 2009; Khatri, Sirota & Butte, 2012; Rhinn et al., 2013; Mitrea et al., 2013). Related to this, a hub gene within a biological network is an important gene which acts to influence the activity of a number of genes (Flintoft, 2004), even influence the activity and function of the entire biological network. Hence, there has been an increased interest to analyze the co-

regulation and co-expression of genes in the biological network, and many approaches have been developed to identify differential co-regulation and co-expression of genes in the subnetwork (Kostka & Spang, 2004; Lai et al., 2004; Reverter et al., 2006; Watson, 2006; Choi & Kendziorski, 2009; Leonardson et al., 2010; Langfelder et al., 2011; Odibat & Reddy, 2012). It has been a trend to extend differential expression analysis to differential network analysis (de la Fuente, 2010).

Most of the approaches for differential network analysis are based on different correlation-based metrics to measure the strength of association between any pairs of nodes in a biological network. Generally, there are three main ways to compare networks for differential network analysis. The first approach handles weighted networks and uses some functions of the edge-specific weight differences as edge weights to construct differential networks (Hudson, Reverter & Dalrymple, 2009; Tesson, Breitling & Jansen, 2010; Liu et al., 2010; Rhinn et al., 2013). The second approach tries to find co-expressed gene sets and identify which correlation patterns are different between sets across conditions (Watson, 2006; Rahmatallah, Emmert-Streib & Glazko, 2014). This approach formulates summary measures that represent co-expression in a biological network and compares the metric between sets. The third approach compares the topology of biological networks across different phenotypes by using measures such as degree of nodes or modularity (Reverter et al., 2006; Zhang et al., 2009).

Although the above methods for differential network analysis can deal with some important biological questions, they are still limited. In general, they are based on a basic hypothesis that some connections between genes across the groups could be thought of as “passenger” events and other connections are unique to either one of groups and thus could be “driver” events that contribute to disease progression and development. Hence, they focus on the contribution of individual differential connections to disease. This

results in several limitations. First, each differential connection is regarded by these methods to have an equal contribution to disease. However, it is well understood that loss of a connection between two hub genes from normal to disease is more deleterious than loss of a connection between two non-hub genes. Second, how differential connections ("driver" connections mentioned above) between pairs of genes are associated with diseases is still not very biologically intuitive, because how the dependency between genes contributes to diseases is usually little understood.

To address these problems, we propose a new PageRank-based method called Pathways of Topological Rank Analysis (PoTRA) to detect pathways associated with cancer. PageRank is an algorithm initially used by Google Search to rank websites in their search engine results (Page et al., 1999). It is a way of measuring the importance of nodes in a network. More generally, PageRank has been applied to other networks, e.g., social networks (Pedroche et al., 2013; Wang et al., 2013). To date, there have been several studies using PageRank for gene expression and network analysis (Morrison et al., 2005; Winter et al., 2012; Kimmel & Visweswaran, 2013; Hou & Ma, 2014; Bourdakou, Athanasiadis & Spyrou, 2016; Zeng et al., 2016; Ramsahai et al., 2017). . These studies focus on ranking genes and discovering key driver genes in disease, and do not try to detect dysregulated pathways involved in disease. Other studies (Winter et al., 2012; Zeng et al., 2016) use PageRank to select topological important genes and simply see which pathways that these topological important genes are involved in. These PageRank-related approaches are very different from our approach.

Our approach embodied by PoTRA is motivated by the observation that the loss of connectivity is a common topological trait of cancer networks (Anglani et al., 2014), as well as the prior knowledge that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs

and a large number of nodes are non-hubs (Albert, 2005; Khanin & Wit, 2006; Zhu, Gerstein & Snyder, 2007). However, from normal to cancer, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network, namely, the number of hub genes might be altered in cancer compared to that in normal samples. Thus, we hypothesize that if the number of hub genes is different in a pathway between normal and cancer, this pathway might be involved in cancer. Based on this hypothesis, we propose to detect pathways involved in cancer by testing if the number of hub genes for each pathway is different between normal and cancer samples. Therefore, the PoTRA approach computes topological ranks of genes in each pathway, and then detects pathways with significantly altered number of hub genes between normal and cancer. Namely, we first use the Google search PageRank algorithm to measure the relative topological ranks of genes in a biological pathway across different conditions. Then, we use Fisher's exact test to estimate if the number of hub genes in each pathway is significantly different between normal and cancer. As an illustration, we apply PoTRA to study hepatocellular carcinoma (HCC) and its subtypes and identify disease-relevant pathways. In conclusion, PoTRA is a new approach to explore and discover cancer-associated pathways. PoTRA can be used as a complement to other existing methods to enrich our understanding of the biological mechanisms behind cancer at the systems-level.

2.2. Materials and Methods

Overview of the PoTRA method

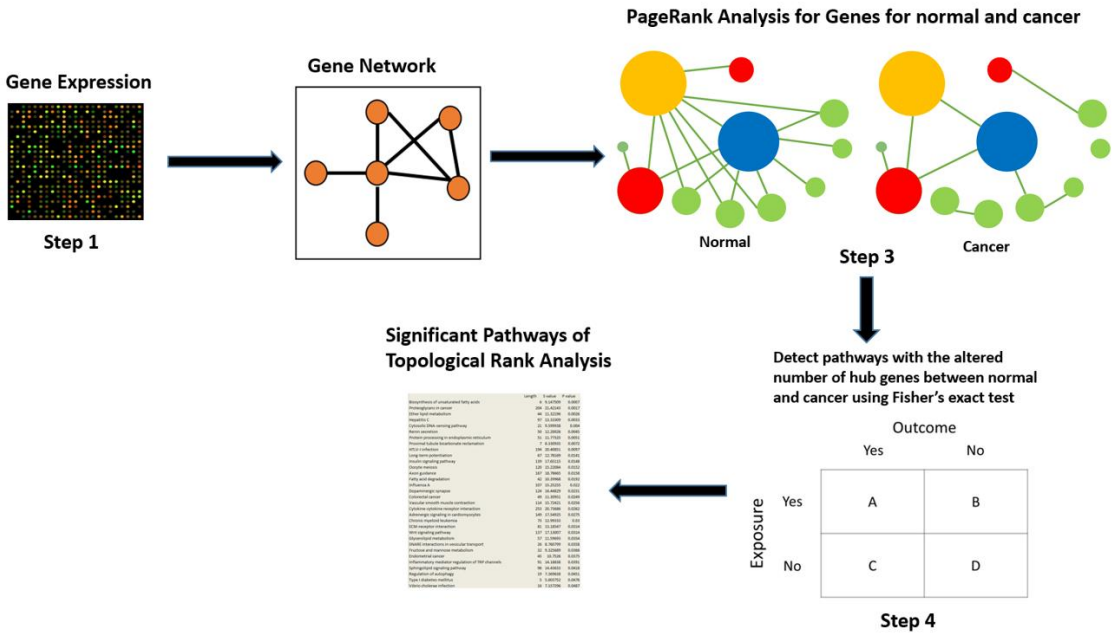


Figure 2.1. Overview of the PoTRA method.

Below we detail the steps of the PoTRA method, as illustrated in **Figure 2.1**.

2.2.1. Data

To illustrate the PoTRA method, we use publicly available gene expression datasets from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) hepatocellular carcinoma (HCC) study. We analyze and contrast 50 HCC samples and 50 tumor-adjacent normal samples ("normal samples" in future sections). In addition, the datasets also include gene expression profiles for several HCC subtypes. We further analyze and contrast 22 hepatitis B-induced HCC samples and 22 tumor-adjacent normal samples, 34 hepatitis C-induced HCC samples and 34 tumor-adjacent normal samples, and 50 alcohol-induced HCC samples and 50 tumor-adjacent normal samples. There are 20,531 gene expression values for each sample. Pathway information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa & Goto, 2000) is used.

To date, there is much known about etiology of HCC (Beasley, 1988; Sanyal, Yoon & Lencioni, 2010; Wang et al., 2012; Goossens & Hoshida, 2015) and knowledge of pathways involved in HCC (Villanueva et al., 2008; Zhou et al., 2010; Wang et al., 2017), which makes it easier to illustrate and assess the PoTRA method.

2.2.2. Construction of gene co-expression network for a pathway

We apply the PoTRA method to gene expression profiles for several phenotypes, such as normal and cancer and cancer subtypes. First, we select genes for each pathway, using pathway information from KEGG. For each pathway, we determine the gene-gene interactions by using the Pearson's correlation to test each co-expressed gene pair. The test calculates a P-value for the association between each pair of genes. A significance threshold of 0.05 is used. When the P-value of a pair of genes is below 0.05, we establish an edge between the corresponding two genes; otherwise, there is no edge between them. We implement it through a built-in function called "cor.test()" in the statistical software package R (https://www.r-project.org/). In this way, we can construct gene co-expression networks (i.e., pathways) for normal and cancer, separately. Of note, all the gene co-expression networks (i.e., pathways) used by PoTRA are undirected graphs, because co-expression networks only focus on gene pairs with a similar expression pattern across samples, in other words, the transcript levels of two co-expressed genes rise and fall together across samples.

2.2.3. PageRank analysis for genes within a pathway for normal and cancer

Based on the above constructed interactions within a pathway, we can observe that some genes are hub genes whereas others are non-hub genes. We apply the PageRank

algorithm (Page et al., 1999) to obtain the corresponding topological importance for each gene within the pathway for normal and cancer, separately, see **Figure 2.2**.

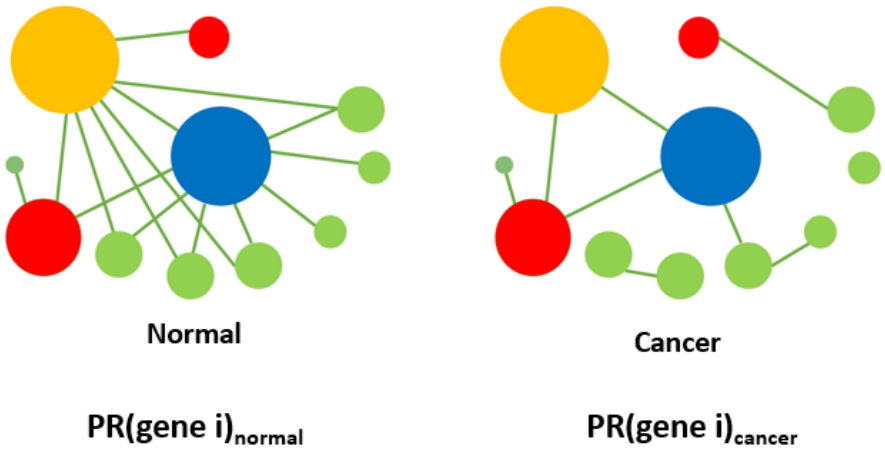


Figure 2.2. The topological rank analysis for each gene within a pathway. For genes within a specified pathway, according to **Step 2**, we construct a corresponding gene co-expression network for normal and cancer, separately. Then we apply the PageRank method to obtain the topological importance of each gene for normal and cancer, separately. PR(gene i)$_{normal}$ represents the PageRank score of the gene i for normal samples, while PR(gene i)$_{cancer}$ represents the PageRank score of the gene i for cancer samples.

We implement it by using the page.rank() function from the igraph (Csárdi & Nepusz, 2006) R package. As mentioned in **Step 2**, all the networks that we construct are undirected graphs. Thus, the PageRank algorithm used in our approach is based on undirected graphs.

2.2.4. Detect pathways with significantly altered number of hub genes between normal and cancer using Fisher's exact test

As mentioned above, PoTRA is motivated by the observation that the loss of connectivity is a common topological trait of cancer networks (Anglani et al., 2014) and the prior

knowledge that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs (Albert, 2005; Khanin & Wit, 2006; Zhu, Gerstein & Snyder, 2007). From normal to cancer, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network, which can result in an altered number of hub genes between normal and cancer. Hence, a statistics that we compare between two phenotypes is the number of hub genes. The PageRank scores of all genes of a pathway form a distribution, and we use the 95th percentile of the distribution (one-tail) in normal samples as cutoff value for hub genes for both normal and cancer samples. The genes in this pathway with PageRank scores that are above the cutoff value are identified as hub genes for this pathway. Then we count the number of hub genes for normal and cancer, separately. Next, we use Fisher's exact test to assess if the number of hub genes is significantly different between normal and cancer. For details, see **Table 2.1**.

	The number of non-hub genes	**The number of hub genes**	**Row total**
Normal	**a**	**b**	**a+b**
Cancer	**c**	**d**	**c+d**
Column total	**a+c**	**b+d**	**a+b+c+d**

Table 2.1. The contingency table for Fisher's exact test. We use the 95th percentile of the distribution (one-tail) in normal samples as cutoff value for hub genes for both normal and cancer samples. The value "a" represents the number of genes whose PageRank scores are below the cutoff value for normal samples. The value "b" represents the number of genes whose PageRank scores are above the cutoff value for normal samples. The values "c" and "d" are the corresponding values for cancer. We use Fisher's exact test to assess if the number of hub genes is significantly different between normal and cancer.

Fisher's exact test estimates the probability of obtaining any such set of values, given by the hypergeometric distribution:

$$\mathbf{P}=\frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}=\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!n!} \qquad (1)$$

Where n=a+b+c+d, and $\binom{i}{j}$ is the binomial coefficient and the symbol "!" indicates the factorial operator.

Formula 1 gives the exact hypergeometric probability of observing this particular arrangement of the data, assuming the given marginal totals, on the null hypothesis that the number of hub genes is the same for a specified pathway between normal and cancer. If this test statistic is significant, it indicates that there is a significantly different number of hub genes between normal and cancer, thereby this pathway might be involved in cancer. By studying many pathways from the KEGG database we generate a multiple hypothesis testing problem. We address this issue by correcting the P-values calculated for each pathway using the False Discovery Rate (FDR) approach, using the Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995)

2.2.5. Software tools

All the analysis is conducted using the R statistical programming language, using the following R Biocoductor packages: graphite for pathway databases, igraph for PageRank function and graph for visualization.

2.3. Results

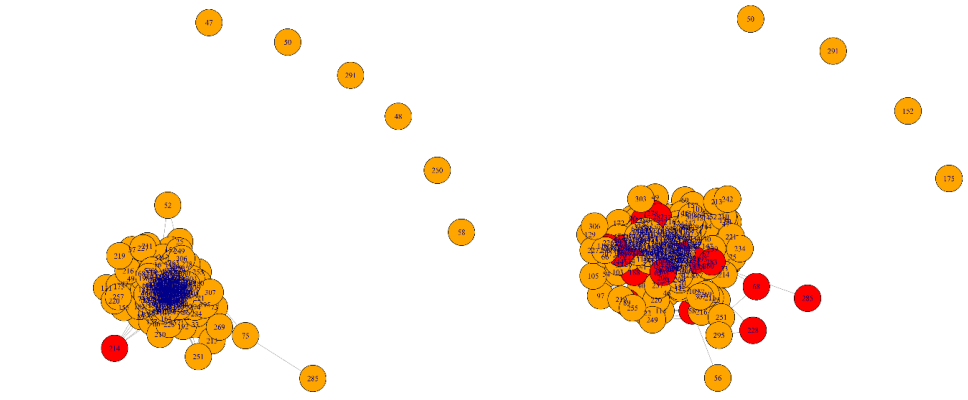
We apply PoTRA to analyze and contrast 50 HCC samples and 50 tumor-adjacent normal samples. All data come from The Cancer Genome Atlas (TCGA) hepatocellular carcinoma (HCC) study.

2.3.1. PoTRA for HCC vs. normal samples

To illustrate the PoTRA method, we use a cancer-associated pathway, "Pathways in cancer", as an example in the following section.

2.3.2. Construction of a gene co-expression network for a pathway

As suggested before, "Pathways in cancer" might be comprised of different interactions between genes under different conditions, such as normal versus cancer conditions. First, we need to find the genes that this pathway consists of by using the KEGG database. In practice, we implement it by using an R package called graphite. Second, for the genes of this pathway, we identify the interactions between genes for normal and cancer samples, separately, see **Figure 2.3**.



a. **b.**

Figure 2.3. Construction of "Pathways in cancer" co-expression network for normal and cancer samples, separately. There are 310 genes in this pathway, labeled 1-310. For the gene names, see **Table 2.7**. Graph (a) represents the network for normal samples and graph (b) represents that for cancer samples. There are 24,924 edges in graph (a) and 9,136 edges in graph (b). The red nodes represent hub genes. There are 16 hub genes in graph (a) and 48 hub genes in graph (b).

Very interestingly, we find a large loss of connectivity in the cancer gene co-expression network with respect to normal ones, from 24,924 to 9,136 edges, which is also noticed by prior literature that has found that the loss of connectivity is a common topological trait of cancer networks (Anglani et al., 2014).

2.3.3. PageRank analysis for genes within a pathway for normal and cancer

Based on the interactions identified in the previous section, and illustrated in **Figure 2.3 (a)** and **(b)**, we can obtain a PageRank score for each gene in "Pathways in cancer" for normal and cancer, separately, which quantifies the influence of a gene on the activity of other genes in this pathway. For the results for this step, 2 vectors with 310 PageRank values, one for normal and one for cancer, separately, see **Table 2.7**. As previously mentioned, the PageRank values in each vector add up to 1.

As an interesting note, in **Figure 2.3**, gene 47 (top of **Figure 2.3(a)**) is an isolated gene in the pathway for normal, but the PageRank score of this gene is not zero, because zero is not allowed for PageRank calculation. Considering this situation, PageRank designs a damping factor p (typically p=0.85) and assigns a small number to this isolated node to solve this issue, for details see (Page et al., 1999).

Figure 2.4 illustrates the distributions of PageRank scores for genes in "Pathways in cancer" for normal and cancer, separately.

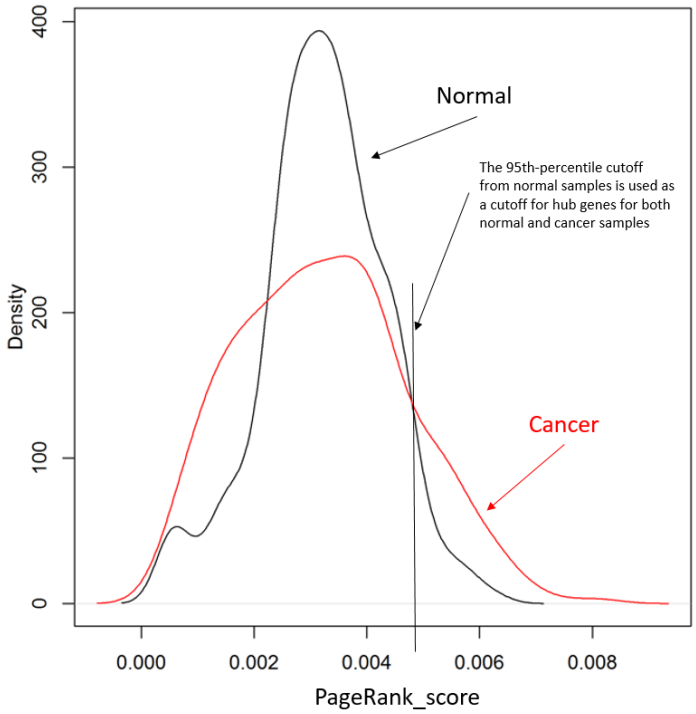


Figure 2.4. The kernel density distribution of PageRank scores of genes in "Pathways in cancer". The red line shows the kernel density distribution of PageRank scores for cancer and the black one is for normal samples. Note that the mean for the two distributions is the same, i.e., mean = 1/N=0.0032258, where N = 310 is the number of genes in the "Pathways in cancer" pathway. We use the 95th-percentile cutoff (= 0.00482) of the kernel distribution in normal samples as cutoff for hub genes for both normal and cancer samples.

Very interestingly, as connectivity is lost in cancer samples, the number of hub genes changes. While there are only 16, strongly-connected (with more edges, an average of 148 edges) hub genes in the normal samples, there are 48 hub genes in the cancer samples, more loosely-connected (with fewer edges, an average of 104 edges).

As mentioned before, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs, namely, the number of hub genes might be altered in cancer compared to that in normal. The altered number of hub genes might be a trait of a pathway in cancer,

which suggests the pathway is involved in cancer if the change in hub gene number is statistically significant.

2.3.4. Fisher's exact test for comparing the number of hub genes in the pathway

We next use Fisher's exact test to test if the number of hub genes for "Pathways in cancer" is significantly different between normal and cancer. The result for the "Pathways in cancer" pathway is included in **Table 2.2**.

	Gene Count(L)	# of edges_normal	# of edges_cancer	# of hub genes_normal	# of hub genes_cancer	Adjusted P-value
Pathways in cancer	310	24924	9136	16	48	0.008

Table 2.2. The "Pathways in cancer" pathway identified by PoTRA for HCC using Fisher's exact test. The P value is adjusted by False Discovery Rate (FDR).

The low P-value in **Table 2.2** indicates that the number of hub genes in cancer samples is significantly different from that in normal samples, suggesting that the "Pathways in cancer" pathway is involved in HCC. This example suggests that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs. Moreover, the loss of connectivity from normal to cancer might lead to disrupting the scale-free structure of the network in cancer, thereby resulting in the fact that the number of hub genes is altered in cancer compared to that in normal.

Then we apply the same approach to other pathways from KEGG to compare HCC vs. normal samples. The significant pathways are shown in **Table 2.3**.

		Gene Count(L)	# of edges_normal	# of edges_cancer	# of hub genes_normal	# of hub genes_cancer	Adjusted P-value
1	Pathways in cancer	310	24924	9136	16	48	0.0081
2	MAPK signaling pathway	252	14005	5170	13	40	0.0158
3	Breast cancer	143	3589	1175	8	29	0.0278

Table 2.3. The significant KEGG pathways identified by PoTRA for HCC using Fisher's exact test. FDR adjusted P-values are below 0.05.

We find three significant pathways with altered number of hub genes between normal and cancer. It is well known that these three pathways are strongly associated with cancer in general. MAPK signaling pathway plays a role in the regulation of gene expression, cellular growth, and survival (Knight & Irving, 2014). Abnormal MAPK signaling might lead to uncontrolled or increased cell proliferation and resistance to apoptosis (Santarpia, Lippman & El-Naggar, 2012; Burotto et al., 2014). Interestingly, we also find that loss of connectivity and the larger number of hub genes for cancer are characteristics of the other two pathways as well.

2.3.5. PoTRA for cancer subtype analysis

Many complex diseases have subtypes and/or can be classified into different categories based on diagnosis, pathology, phenotype characteristics, etc. To further assess the PoTRA method, we apply it to several subtypes of the HCC TCGA data. There are several risk factors associated with HCC, such as hepatitis B, hepatitis C and alcohol (Beasley, 1988; Sanyal, Yoon & Lencioni, 2010; Hoshida et al., 2014; Goossens & Hoshida, 2015).

Here, we apply PoTRA to compare these three subtypes of HCC samples with normal samples.

Table 2.4 illustrates the Fisher's exact test results for comparing normal with hepatitis B-induced HCC samples.

		Gene Count(L)	# of edges_ normal	# of edges_ cancer	# of hub genes_ normal	# of hub genes_ cancer	Adjusted P-value
1	Insulin signaling pathway	139	2692	958	7	34	0.0007
2	Pathways in cancer	310	11194	3792	16	52	0.0007
3	Hippo signaling pathway	151	2836	970	8	31	0.0072
4	HTLV-I infection	194	5518	2080	10	35	0.0072
5	Neurotrophin signaling pathway	117	2441	895	6	25	0.0195
6	mTOR signaling pathway	144	3410	832	8	28	0.0240
7	Epstein-Barr virus infection	85	1524	435	5	21	0.0353
8	Hepatitis B	134	2708	828	7	25	0.0353

Table 2.4. The significant KEGG pathways identified by PoTRA for hepatitis B-induced HCC using Fisher's exact test. FDR adjusted P-values are below 0.05.

There is one common pathway, Pathways in cancer, between **Table 2.4** and **Table 2.3**. There are seven other new pathways, which are very interesting and associated with the hepatitis B-induced HCC. First, the "Hepatitis B" pathway is detected by our method. Hepatitis B is the most important and direct factor causing hepatitis B-induced HCC. In addition, we find two other pathways, HTLV-I (Human T-cell leukemia virus type I) infection and Epstein-Barr virus infection, which are strongly associated with virus

infection and cancer. This is consistent with the viral pathology of hepatitis B-induced HCC. Besides, some studies show that hepatitis B virus infection can contribute to the impairment of insulin signaling, which is another pathway identified by PoTRA (Kim, Kim & Cheong, 2010; Barthel et al., 2016). Finally, the other three pathways, Hippo signaling pathway, Neurotrophin signaling pathway and mTOR signaling pathway are associated with cancer in general. Hippo signaling pathway is reported to be able to control organ size through regulating cell proliferation and apoptosis (Saucedo & Edgar, 2007; Pan, 2010). It is reported that neurotrophins can regulate cancer stem cells (Chopin et al., 2016), and neurotrophins contribute to pro-survival signaling in many different types of cancer (Molloy, Read & Gorman, 2011). mTOR signaling pathway is a well-known cancer-associated pathway. Alterations of mTOR signaling pathway have significant effects on cancer progression. The major components of mTOR signaling pathway are critical effectors in cell signaling pathways commonly deregulated in cancers (Guertin & Sabatini, 2007; Villanueva et al., 2008; Pópulo, Lopes & Soares, 2012).

These results suggest that PoTRA can be used to identify not only the pathways associated with cancer in general, but also those pathways associated with cancer subtypes, such as hepatitis B-induced HCC specifically.

Results of the PoTRA analysis from two other HCC subtypes, hepatitis C-induced HCC and alcohol-induced HCC, are included in **Table 2.5** and **Table 2.6**, respectively.

		Gene Count(L)	# of edges_ normal	# of edges_ cancer	# of hub genes_ normal	# of hub genes_ cancer	Adjusted P-value
1	Pathways in cancer	310	22253	7791	16	62	2.89E-06

2	PI3K-Akt signaling pathway	340	19901	6594	17	65	2.89E-06
3	MAPK signaling pathway	252	11986	4168	13	47	0.0003
4	Proteoglycans in cancer	204	9815	3642	11	38	0.0033
5	Rap1 signaling pathway	208	8294	3587	11	34	0.0215
6	Adrenergic signaling in cardiomyocytes	149	3594	1355	8	27	0.0372
7	cAMP signaling pathway	196	5106	2493	10	30	0.0372
8	Focal adhesion	203	10225	4656	11	32	0.0372
9	HTLV-I infection	194	9843	4030	10	30	0.0372
10	Ras signaling pathway	226	10098	3931	12	33	0.0376
11	FoxO signaling pathway	126	3391	1222	7	24	0.0380
12	Osteoclast differentiation	123	4418	1452	7	24	0.0380
13	ErbB signaling pathway	88	2128	814	5	20	0.0400
14	Axon guidance	167	6203	2705	9	27	0.0433

Table 2.5. The significant KEGG pathways identified by PoTRA for hepatitis C-induced HCC using Fisher's exact test. FDR adjusted P-values are below 0.05.

In **Table 2.5**, we find two common pathways, Pathways in cancer and MAPK signaling pathway, between **Table 2.5** and **Table 2.3**. Among the other pathways, we find several pathways related to cancer generally and hepatitis C-induced HCC specifically. First, HTLV-I infection is also listed in this table, and, as mentioned above, is associated with virus infection and cancer. Almost all other pathways are associated with cancer in general. PI3K-Akt signaling pathway is a key regulator of normal cellular processes involved in cell growth, proliferation, motility, survival, and apoptosis (Porta, Paglino & Mosca, 2014). The Proteoglycans in cancer pathway is involved in regulation of

proteoglycans, heavily glycosylated proteins present especially in connective tissue in cancer. Rap1 signaling pathway is reported to be involved in cancer cell migration, invasion and metastasis (Bailey, 2009; Zhang et al., 2017). The cAMP signaling pathway regulates a number of biological processes, such as cell growth and adhesion, neuronal signaling, energy homeostasis and muscle relaxation (Fajardo, Piazza & Tinsley, 2014). The key component of Focal adhesion pathway, Focal adhesion kinase (FAK), is reported to enable activation by growth factor receptors or integrins in different types of cancers. FAK is an important mediator of cell proliferation, cell migration, cell growth (Golubovskaya, Kweh & Cance, 2009; Tai, Chen & Shen, 2015). A large volume of literature shows Ras signaling pathway is involved in several aspects of normal cell growth and malignant transformation, and plays an important role in cancer development and progression (Vojtek & Der, 1998; Downward, 2003; Santarpia, Lippman & El-Naggar, 2012; Knight & Irving, 2014). FoxO signaling pathway is involved in the regulation of the cell cycle, apoptosis and metabolism (Schmidt et al., 2002; Fu & Tindall, 2008; Gross, van den Heuvel & Birnbaum, 2008). Besides, activity of FoxO signaling pathway also affects stem cell maintenance and lifespan (Eijkelenboom & Burgering, 2013). ErbB signaling pathway plays roles in cancer development and progression (Hynes & Lane, 2005; Seshacharyulu et al., 2012), as well as in cancer cell migration and invasion (Appert-Collin et al., 2015). ErbB signaling pathway is associated with the development of a wide variety of types of solid tumor if ErbB signaling is excessive (Cho & Leahy, 2002). The Axon guidance pathway is also reported to regulate cell migration and apoptosis, and be associated with tumorigenesis (Chédotal, Kerjan & Moreau-Fauvarque, 2005).

		Gene Count(L)	# of edges_ normal	# of edges_ cancer	# of hub genes_ normal	# of hub genes_ cancer	Adjusted P-value
1	PI3K-Akt signaling pathway	340	23928	8733	17	55	0.0006
2	MAPK signaling pathway	252	14005	5767	13	46	0.0007
3	Pathways in cancer	310	24924	10191	16	47	0.0043

Table 2.6. The significant KEGG pathways identified by PoTRA for alcohol-induced HCC using Fisher's exact test. FDR adjusted P-values are below 0.05.

We find two common pathways between **Table 2.6** and **Table 2.3**, MAPK signaling pathway and Pathways in cancer. As mentioned above, PI3K-Akt signaling pathway also plays an important role in cancer (Porta, Paglino & Mosca, 2014).

2.4. Discussion

We propose a PageRank-based method, Pathway of Topological Rank Analysis (PoTRA), for identifying pathways involved in cancer. PoTRA is motivated by the observation that the loss of connectivity is a common topological trait of cancer networks (Anglani et al., 2014) and the prior knowledge that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs

and a large number of nodes are non-hubs (Albert, 2005; Khanin & Wit, 2006; Zhu, Gerstein & Snyder, 2007). From normal to cancer, the process of the network losing connectivity might be the process of disrupting the scale-free structure of the network, which can result in an altered number of hub genes between normal and cancer. The PoTRA analysis is based on topological ranks of genes in biological pathways, and PoTRA detects pathways involved in cancer by testing if the number of hub genes in pathways is altered between normal and cancer.

To illustrate the method, PoTRA is applied to several TCGA hepatocellular carcinoma datasets. The results in our study are in agreement with prior knowledge of HCC from literature. We find that a high proportion of statistically significant pathways play important roles in cancer, indicating that the altered number of hub genes for these pathways might indeed be a reflection of the underlying biological causes that lead to cancer. Moreover, in the comparison between normal and each subtype of HCC, most importantly, the “Hepatitis B” pathway and several pathways associated with virus infection dramatically become significant pathways in hepatitis B-induced HCC, suggesting that PoTRA is capable of detecting pathways associated with disease subtypes. We also find several pathways associated with HCC generally and subtype specifically in hepatitis C-induced HCC and in alcohol-induced HCC.

In our approach, the correlation method is used to construct gene co-expression networks for normal and cancer, respectively. A gene co-expression network is an undirected graph, where each node represents a gene, and each edge is established if there is a significant co-expression relationship between two genes (Stuart et al., 2003). Stuart JM, et al. (Stuart et al., 2003) found 22,163 co-expression relationships, each of which has been conserved across evolution, suggesting that the co-expressions between genes confers a selective advantage and thus these genes are functionally related. Gene

co-expression networks are biologically interesting since co-expressed genes might be controlled by members of the same pathway, or the same transcriptional regulatory program or protein complex (Weirauch, 2011), and could be functionally related, suggesting that co-expression is common in the human genome. A gene co-expression network can be constructed by looking for pairs of genes with a similar expression pattern across samples, i.e., the transcript levels of two co-expressed genes rise and fall together across samples. Therefore, the correlation method used by PoTRA is capable to identify co-expressed gene pairs.

2.5. Future Directions

The hypothesis of our study is based on the fact that the loss of connectivity is a common topological trait of cancer networks (Anglani et al., 2014). It is not yet well understood if this trait is a characteristic of other complex diseases. Thus, we need to be cautious about the applicability of this method to other diseases. However, this trait could be applicable to other complex diseases. Thus, although PoTRA is motivated by work on cancer, it could apply to other complex diseases as well. This area needs to be further investigated.

In this study, we apply PoTRA to pre-defined biological pathways, from the well-curated KEGG pathway database. However, the PoTRA method can also be applied to any set of genes of interest, such as functional gene subnetworks. This could be an interesting area to further explore.

In addition to hub genes, bottleneck genes which have a high betweenness centrality (i.e., network nodes that have many “shortest paths” going through them, analogous to major bridges and tunnels on a highway map) are also fairly important for a biological network. Bottleneck gene are key connectors in gene network, bottlenecks are shown to

tend to be essential in gene network. Therefore, bottlenecks might be a good direction to further investigate.

We can further validate the PoTRA method using simulated data. We might simulate a network (i.e., biological pathways) composed of 100 nodes (i.e., genes) using Barabasi-Albert model (Albert & Barabási, 2002). Also, we simulate an Erdos-Renyi random network (Erdős, 1959) composed of 100 nodes using an Erdos-Renyi random graph generator embedded in the R package "igraph" (Csárdi & Nepusz, 2006). The former network is a scale-free network, while the latter network is a random graph. Hence, the expected result is that those two networks are significantly different. Then PoTRA is applied to test if those two networks are different.

For the PageRank algorithm, we use the default value of damping factor 0.85 in this study. However, the damping factor is not necessarily a fixed value in different cases. Hence, we might further use simulation to decide the best value of damping factor in different cases. This is a very interesting and valuable direction to explore.

For those significant pathways that we identify, we might further hypothesize that there are some common genes in those significant pathways. These common genes could be driver genes to drive tumorigenesis. We might further find those common genes and investigate the biology behind some of results.

2.6. Conclusion

In summary, PoTRA provides a new method for detecting cancer-associated pathways. PoTRA may be used to augment existing methods and provide a richer, more systematic understanding of cancer mechanisms.

2.7. Availability

R software to carry out the PoTRA computation is available via

http://www.dinulab.org/tools.

2.8. Supplementary Materials

	Gene Symbol	PR_normal	PR_cancer		Gene Symbol	PR_normal	PR_cancer
1	AKT3	0.003093905	0.00300372	156	MAP2K2	0.004305054	0.001979433
2	CDK2	0.003040914	0.003949561	157	BAD	0.002098977	0.001537706
3	CDK4	0.003391899	0.003442728	158	PTCH1	0.004896564	0.005416168
4	CDK6	0.002710498	0.002632461	159	PTEN	0.003402329	0.004924039
5	CDKN1A	0.003458217	0.001332768	160	PTK2	0.002923717	0.004069311
6	CDKN1B	0.003636721	0.004189972	161	BAX	0.002922207	0.002322378
7	CDKN2A	0.003618423	0.00271539	162	RAC1	0.002793632	0.004790566
8	CDKN2B	0.004020755	0.007139137	163	RAC2	0.003750833	0.003195001
9	LAMC3	0.003079645	0.002038664	164	RAC3	0.00343216	0.002457697
10	TFG	0.003411381	0.003289014	165	RAF1	0.003089542	0.005300568
11	CEBPA	0.002461461	0.003044095	166	RALA	0.002535952	0.005790368
12	RALBP1	0.002763423	0.005092257	167	RALB	0.003030614	0.006514099
13	RASSF1	0.002415298	0.003832637	168	RALGDS	0.002249478	0.00546576
14	FZD10	0.005169905	0.001816191	169	RARA	0.005698338	0.004808868
15	EGLN2	0.00353983	0.001187428	170	RARB	0.002455165	0.002768134
16	EGLN3	0.003632638	0.002688535	171	RB1	0.002302646	0.005971465
17	CHUK	0.003863949	0.004152228	172	CCND1	0.004082443	0.001725468
18	CKS1B	0.004536644	0.002561939	173	BCL2	0.002264799	0.003741892

19	CKS2	0.00403055	0.00214326	174	RELA	0.00362278	0.003817114
20	COL4A1	0.002642866	0.00282523	175	RET	0.001689008	0.000489237
21	COL4A2	0.003103248	0.002153239	176	BCR	0.004432436	0.004640876
22	COL4A3	0.003271212	0.001209817	177	RXRA	0.002619723	0.001323734
23	COL4A4	0.002970752	0.003097248	178	RXRB	0.002878814	0.001746663
24	COL4A5	0.006294353	0.001287653	179	BID	0.003641528	0.002462616
25	COL4A6	0.001253758	0.000767639	180	SHH	0.004395602	0.004216858
26	CRK	0.003741232	0.005513751	181	SKP2	0.003184264	0.002789141
27	CRKL	0.00325282	0.004265987	182	SMO	0.002328005	0.00409342
28	CTBP1	0.003456445	0.006251871	183	SOS1	0.003116838	0.003698003
29	CTBP2	0.002354344	0.003178328	184	SOS2	0.003323926	0.005023092
30	CTNNB1	0.002542315	0.003788528	185	SPI1	0.00434689	0.003165557
31	DAPK1	0.003198525	0.005002724	186	BRAF	0.003482398	0.004239558
32	DAPK3	0.004720648	0.004015498	187	STAT1	0.003735686	0.003035884
33	DCC	0.001960459	0.001561484	188	STAT3	0.004346367	0.00316925
34	DVL1	0.004021144	0.003689196	189	STAT5A	0.003173093	0.005400298
35	DVL2	0.002631472	0.003972589	190	STAT5B	0.003293719	0.003781533
36	DVL3	0.002227315	0.003796052	191	TCEB1	0.004751576	0.003317497
37	EGF	0.00151036	0.002656015	192	TCEB2	0.001395063	0.003835255
38	EGFR	0.003054758	0.005109083	193	TCF7	0.002361603	0.003532294
39	EPAS1	0.003394586	0.004924966	194	TCF7L2	0.003775016	0.003724009
40	ERBB2	0.004808457	0.001814885	195	TGFA	0.003914048	0.003407188
41	AKT1	0.004401951	0.001514189	196	TGFB1	0.003614268	0.003182874
42	AKT2	0.003416283	0.004092489	197	TGFB2	0.005255924	0.004529128
43	ETS1	0.002538532	0.002663598	198	TGFB3	0.003103419	0.003000592
44	MECOM	0.003437083	0.004944226	199	TGFBR1	0.00225014	0.003598152
45	FGF1	0.003430158	0.001526164	200	TGFBR2	0.003613977	0.006812136
46	FGF2	0.004045796	0.001289223	201	TP53	0.00252106	0.002459

47	FGF3	0.000491965	0.001395881	202	TPM3	0.002578494	0.004063351
48	FGF4	0.000491965	0.001067975	203	TPR	0.002005053	0.003321939
49	FGF5	0.002047965	0.001584524	204	HSP90B1	0.004129877	0.003999821
50	FGF6	0.000491965	0.000489237	205	VHL	0.003158384	0.004027472
51	FGF7	0.002577353	0.003851117	206	WNT1	0.003694857	0.00230586
52	FGF8	0.000606536	0.001337488	207	WNT2	0.003875735	0.001715735
53	FGF9	0.004940821	0.005184195	208	WNT3	0.004378124	0.002437254
54	FGF10	0.002788004	0.002202358	209	WNT5A	0.003303738	0.002342169
55	FGF11	0.003096608	0.004479963	210	WNT6	0.001516543	0.001345974
56	FGF12	0.002897905	0.001349453	211	WNT7A	0.001472326	0.001254448
57	FGF13	0.004025308	0.001043084	212	WNT7B	0.002526824	0.00251065
58	FGF14	0.000491965	0.001825604	213	WNT8A	0.000861505	0.000595714
59	FGFR1	0.003588912	0.00272812	214	WNT8B	0.000732661	0.000732088
60	FGFR3	0.004824156	0.002028192	215	WNT10B	0.003746589	0.001971191
61	FGFR2	0.002679216	0.002991078	216	WNT11	0.001458332	0.001480431
62	LAMB4	0.003877415	0.001990161	217	WNT2B	0.004452958	0.002412174
63	FLT3	0.0045477	0.00406583	218	WNT9A	0.004818818	0.002510426
64	FLT3LG	0.004221203	0.002718439	219	WNT9B	0.000764811	0.00141332
65	FN1	0.002612425	0.004331963	220	ZBTB16	0.001507095	0.00151374
66	FOS	0.002060237	0.002313126	221	PAX8	0.004644468	0.001018703
67	PIK3R5	0.003821391	0.003815514	222	FZD5	0.00309711	0.002893059
68	DAPK2	0.003393029	0.000754584	223	FZD3	0.00419619	0.002834875
69	CBLC	0.00419537	0.001124637	224	CCDC6	0.003320029	0.00622032
70	ABL1	0.002828823	0.0054705	225	NCOA4	0.002797625	0.004216662
71	FZD2	0.002945035	0.003247979	226	WNT10A	0.004112892	0.003183438
72	APPL1	0.003033474	0.004534701	227	FGF23	0.001459143	0.001574117
73	FGF20	0.001614757	0.001404174	228	WNT5B	0.002563388	0.000961552
74	FGF21	0.003062587	0.003391477	229	FZD1	0.002984098	0.003891502

75	FGF22	0.001146482	0.000975682	230	FZD4	0.003876819	0.002179775
76	STK36	0.002642641	0.003945146	231	FZD6	0.003020781	0.004240562
77	GLI1	0.003813176	0.002437481	232	FZD7	0.003278225	0.002572302
78	GLI2	0.003033217	0.002735696	233	FZD8	0.004655819	0.00337092
79	GLI3	0.002741522	0.002361796	234	FZD9	0.002771345	0.000786142
80	LAMA1	0.005807073	0.001932564	235	TCF7L1	0.004644479	0.00349625
81	GRB2	0.002216502	0.005585664	236	RASSF5	0.002955222	0.002820201
82	GSK3B	0.002274372	0.00450013	237	CASP3	0.003064809	0.003100381
83	HDAC1	0.002668862	0.004121253	238	CASP8	0.003093951	0.005910365
84	HGF	0.003404592	0.002564872	239	CASP9	0.005074869	0.001516088
85	HIF1A	0.003504699	0.005674787	240	CUL2	0.002964537	0.004349557
86	HRAS	0.005395857	0.002094702	241	PIK3R3	0.003215646	0.006488887
87	HSP90AA1	0.00370572	0.005317323	242	IKBKG	0.004975669	0.001201659
88	HSP90AB1	0.0026485	0.003443134	243	RUNX1	0.004289037	0.004484772
89	IGF1	0.00189462	0.001526586	244	RUNX1T1	0.003926335	0.001790019
90	IGF1R	0.003374734	0.000692951	245	CBL	0.003308481	0.005999363
91	FAS	0.002947965	0.002274438	246	CBLB	0.00236811	0.008074871
92	IKBKB	0.002621242	0.005865599	247	FADD	0.004632577	0.004130511
93	FASLG	0.003489903	0.003679946	248	FGF18	0.005343563	0.002291649
94	ITGA6	0.002667856	0.003401315	249	FGF17	0.001685853	0.001240543
95	AR	0.001843612	0.003851707	250	FGF16	0.000491965	0.001716323
96	ITGA2	0.004255997	0.003790406	251	WNT3A	0.000670711	0.001163439
97	ITGA2B	0.001779879	0.001293375	252	CCNE1	0.003777607	0.004063126
98	ITGA3	0.004571272	0.002333706	253	PIAS2	0.003294104	0.003706514
99	ITGAV	0.002577937	0.004511556	254	CCNE2	0.003829759	0.003024707
100	ITGB1	0.002722552	0.005512016	255	FGF19	0.002041663	0.001547215
101	ARAF	0.003770923	0.001648577	256	RBX1	0.003486231	0.00197113
102	JAK1	0.00215008	0.006380458	257	FOXO1	0.001463804	0.001110548

103	JUN	0.004604203	0.001826207	258	MTOR	0.002418695	0.002870981
104	JUP	0.004523251	0.003669617	259	CSF3R	0.004223332	0.003860835
105	KIT	0.004796668	0.000915436	260	E2F1	0.003990438	0.003854594
106	KRAS	0.003798401	0.005525741	261	E2F2	0.004058837	0.0054384
107	RHOA	0.002574591	0.004360737	262	E2F3	0.004336313	0.005194975
108	LAMA2	0.003538299	0.001851764	263	IL6	0.002959906	0.003059989
109	LAMA3	0.004591861	0.000850307	264	CDC42	0.002459057	0.00483741
110	LAMA4	0.002821443	0.003242991	265	STK4	0.002657375	0.005353271
111	LAMA5	0.005952102	0.005213777	266	HDAC2	0.003097869	0.004292515
112	LAMB1	0.002557073	0.004371683	267	CREBBP	0.002935804	0.003175403
113	LAMB2	0.003725664	0.003748125	268	EP300	0.00263447	0.005360202
114	LAMB3	0.004672454	0.000787888	269	VEGFD	0.001163632	0.000784671
115	LAMC1	0.002311235	0.00289013	270	ARNT	0.002580086	0.004391662
116	LAMC2	0.002608833	0.00198952	271	PGF	0.004505714	0.002935848
117	SMAD2	0.002249808	0.004860795	272	SLC2A1	0.005452944	0.003071782
118	SMAD3	0.002707199	0.004199273	273	VEGFA	0.0039616	0.004470881
119	MAX	0.002868878	0.005076557	274	VEGFB	0.002820038	0.003698483
120	MDM2	0.003558278	0.00347452	275	VEGFC	0.003742468	0.003304904
121	MET	0.003552921	0.004114956	276	ARNT2	0.003321398	0.002836696
122	KITLG	0.004774841	0.002835445	277	CXCL8	0.004633571	0.003557386
123	MYC	0.004252456	0.002505314	278	MMP1	0.002590167	0.002904235
124	NFKB1	0.004045736	0.00458389	279	MMP2	0.003286815	0.002153179
125	NFKB2	0.003607195	0.004612478	280	MMP9	0.002602466	0.001893486
126	NFKBIA	0.004361484	0.001949027	281	HHIP	0.003682677	0.002921235
127	NKX3-1	0.004297668	0.002708324	282	BMP2	0.003422092	0.002578134
128	NRAS	0.002567488	0.00432757	283	BMP4	0.003690261	0.002081375
129	NTRK1	0.004454522	0.001981356	284	PTCH2	0.002954986	0.003976521
130	LEF1	0.003495041	0.002350695	285	KLK3	0.000509865	0.000737893

131	WNT16	0.001098482	0.001409997	286	PPARD	0.00302206	0.005249179
132	PDGFA	0.003389003	0.002862777	287	MAPK8	0.003270075	0.003491051
133	PDGFB	0.004505859	0.004401392	288	MAPK9	0.002383262	0.003928425
134	PDGFRA	0.002673138	0.003186629	289	MAPK10	0.002947271	0.002123566
135	PDGFRB	0.00297555	0.003415683	290	SMAD4	0.003325315	0.004625337
136	SUFU	0.002317984	0.00358168	291	BIRC8	0.000491965	0.000489237
137	PIK3CA	0.003414598	0.004320752	292	BIRC2	0.002887534	0.003677406
138	PIK3CB	0.002565891	0.004433264	293	BIRC3	0.004145515	0.002910634
139	PIK3CD	0.00347652	0.002307271	294	XIAP	0.003063954	0.005127472
140	PIK3CG	0.003017476	0.003016939	295	NOS2	0.0020443	0.000828748
141	PIK3R1	0.00334452	0.003301249	296	PTGS2	0.005697117	0.002183857
142	PIK3R2	0.004332678	0.002937172	297	BCL2L1	0.004366514	0.003791546
143	PLCG1	0.002311099	0.00631182	298	TRAF1	0.003487191	0.003371306
144	PLCG2	0.003709433	0.002335372	299	TRAF2	0.003406522	0.003587955
145	PML	0.003385369	0.00523318	300	TRAF3	0.002531222	0.004840198
146	CYCS	0.004858868	0.003588693	301	TRAF5	0.003799039	0.004325504
147	WNT4	0.004326655	0.001893445	302	TRAF6	0.003200779	0.006150486
148	EGLN1	0.004057612	0.004247422	303	BIRC7	0.003638758	0.001685454
149	PPARG	0.004018764	0.003434064	304	TRAF4	0.003163841	0.002732923
150	PRKCA	0.003893901	0.00340499	305	BIRC5	0.004552145	0.002196816
151	PRKCB	0.00300293	0.002937265	306	CCNA1	0.003085372	0.001829558
152	PRKCG	0.0019237	0.000489237	307	RXRG	0.00263539	0.002051522
153	MAPK1	0.002567474	0.003750868	308	PLD1	0.003675508	0.002758496
154	MAPK3	0.002835883	0.00451274	309	CSF1R	0.004301397	0.003872604
155	MAP2K1	0.001598758	0.005993616	310	CSF2RA	0.004469274	0.005738579

Table 2.7. The PageRank scores of genes in "Pathways in cancer" for normal and cancer. We consolidate the information from **Section 1.1** – **Section 1.2** into one table with four columns, namely: (1) Gene index (1-310) (2) Gene symbol (3) The PageRank score of each gene in this pathway for normal (4) The PageRank score of each gene in this pathway for cancer.

Chapter 3: miR2Pathway: A Novel Analytical Method to Discover MicroRNA-mediated Dysregulated Pathways Involved in Disease

3.1. Introduction

MicroRNAs (miRNAs) are short non-coding RNAs of about 22 nucleotides in length, involved in the post-transcriptional regulation of gene expression. MiRNAs induce the degradation of mRNA or translational repression of mRNA depending on the degree of homology to specific sequences, typically in the untranslated regions (UTRs) of their targets (Pasquinelli, 2012). MiRNAs are able to impact the expression of one or many genes at a time. It is believed that more than 60% of human genes are regulated by miRNAs (Friedman et al., 2009). Hence, miRNAs can be important regulators of biological processes. For example, a single miRNA can control a complex biological pathway by simultaneously targeting multiple mRNAs of this specific biological pathway, and these targeted mRNAs might be members of a cascade functioning towards a functional endpoint in the same biological pathways or in the crosstalk between biological pathways (Lima et al., 2011). Over the past years, the role of miRNAs as key regulators that control a wide variety of fundamental biological processes involved in proliferation, apoptosis, cell growth, differentiation, invasiveness, motility, other oncogenic related processes, etc., has been demonstrated (Calin & Croce, 2006; Kefas et al., 2008; Saydam et al., 2009; Ponomarev et al., 2011; van Kouwenhove, Kedde & Agami, 2011; Lu et al., 2011; Glass & Singla, 2011; Fu et al., 2012; Fang et al., 2012). For example, Lu and colleagues show that miRNA-21 is able to down-regulate the activity of the IL-12/IFN-γ pathway in lung cancer (Lu et al., 2011, p.). Other research teams have found that miRNA-7 can simultaneously target multiple genes of the PI3-kinase/Akt pathway in hepatocellular carcinoma and glioblastoma (Kefas et al., 2008; Fang et al.,

2012). MiRNA-200 functions as a multifunctional tumor suppressor in meningiomas through multiple and simultaneous influences on the E-cadherin and Wnt/β-catenin signaling pathways (Saydam et al., 2009). MiRNA-106a has been shown to directly inhibit ULK1 mRNA expression levels in acute myeloid leukemia (AML) cells, and also can target other members of the ULK1 complex, such FIP200 and mAtg13 (Fu et al., 2012). The C/EBP-α-PU.1 pathway is found to be regulated by miRNA-124 (Ponomarev et al., 2011), and miRNA-1 has been suggested to inhibit Pten/Akt pathway (Glass & Singla, 2011). MiRNAs are often aberrantly expressed in tumor tissue even in early stages of tumor and other conditions (Negrini et al., 2007; Croce, 2009), which can make them valuable biomarker candidates, such as for Alzheimer's disease (AD) (Leidinger et al., 2013). Furthermore, several studies have demonstrated a potential value of miRNA-based therapy in cancer (Takamizawa et al., 2004; Cimmino et al., 2005; Scott et al., 2007; Kasinski & Slack, 2011). A good example is the utility of anti-miRNA-21 in breast cancer, resulting in suppression of tumor growth in vivo and cell growth in vitro (Si et al., 2007). Therefore, miRNAs' potential as disease biomarkers and therapeutic agents places this group of small non-coding RNAs at the cutting-edge position of biomedical interest. Therefore, an important question is raised based on the generality of the above phenomena: What degree of miRNA-mediated dysregulation of biological pathways is present in disease?

Over the past decade, there has been a large volume of literature demonstrating that miRNAs dysregulate mRNA expression levels by their aberrant expression in diseases (Takamizawa et al., 2004; Cimmino et al., 2005; Scott et al., 2007; Si et al., 2007; Kefas et al., 2008; Saydam et al., 2009; Ponomarev et al., 2011; Lu et al., 2011; Glass & Singla, 2011; Fang et al., 2012; Leidinger et al., 2013; Ding et al., 2015). That is to say, the targeted mRNAs' expression might be aberrantly altered because they are incorrectly

regulated by aberrantly expressed miRNAs. Based on this, several tools have been developed to detect miRNA-pathway associations (Nam et al., 2009; Maragkakis et al., 2011; Hsu et al., 2011a; Lu et al., 2012; Ben-Hamo & Efroni, 2015; Godard & van Eyll, 2015; Preusse, Theis & Mueller, 2016; Backes et al., 2016; Han et al., 2016; Backes et al., 2017). These tools typically propose enrichment-based methods to study associations between miRNAs and pathways. These enrichment-based methods, however, have two common limitations. First, they study the association between miRNA and pathway based on enrichment analysis of targeted genes in a pathway. Hence, they ignore the topological importance of targeted genes in a pathway. For example, if a miRNA targets genes of topological importance in a pathway, such as hub genes, this miRNA might have higher association with the activity of this pathway than another miRNA that targets less topologically important genes. Second, these methods do not aim to identify changes in pathways between disease and non-disease. Instead, they focus on how miRNAs' aberrant expression affect the targeted genes in a pathway, and do not assess how rewired miRNA-mRNA connections influence a pathway.

It is well known that genes interact in complex networks that govern cellular processes. Researchers have discovered how rewired miRNA-mRNA connections influence biological processes in cancer. An important reason why the rewired miRNA-mRNA connections influence biological processes in cancer is that miRNA-mRNA connections tend to be dynamic or condition-specific, or differential between disease and non-disease. For example, Volinia and colleagues (Volinia et al., 2010) have analyzed the genetic networks of miRNAs in cancer, and suggested that in normal tissues, miRNAs are connected in networks and different cell types have different network connections. In cancer, they suggest that it is likely that normal network connections become disrupted or rewired, which might contribute to disease. In addition, Chen-Ching Lin, et al. (Lin et

al., 2015) have identified a regulatory feedback loop between STAT1 and miRNA-155-5p that is consistently activated in cancer, and found that the rewired regulatory networks are highly associated with cancer. Sivan Elhanati, et al. (Elhanati et al., 2016) have found that miRNA-122 and SIRT6 negatively regulate each other's expression, and the connection between them is manifested in two physiologically relevant ways in the liver. First, they negatively regulate a similar set of metabolic genes and fatty acid β-oxidation. Second, they found that the loss of a negative correlation between SIRT6 and miRNA-122 expression is significantly associated with better prognosis in hepatocellular carcinoma patients. There are also analytical approaches for exosomal miRNA expression analysis (Aqil et al., 2014, 2015). Thus, there is an increasing number of relevant studies suggesting that rewired connections between miRNAs and genes are associated with diseases. However, these studies mainly focus on rewired connections between miRNAs and genes, but do not discuss how those rewired miRNA-mRNA connections are associated with dysregulation of biological pathways at the pathway-level.

A recent methodology developed by Kang et al can analyze topological features of miRNA-target gene differential regulatory network (Kang et al., 2017). However, they use "degree" as the topological measurement in their study. "Degree" does not consider the topological weight of each gene in gene regulatory networks. Hence, we need to consider this in our study. Our analysis uses PageRank, an algorithm initially used by Google Search to rank websites in search engine results (Page et al., 1999). It is a way of measuring the importance of nodes in a network. More generally, PageRank has been applied to other networks, e.g., social networks (Pedroche et al., 2013; Wang et al., 2013). To date, there have been several studies that also use PageRank to analyze miRNA (Noh et al., 2014; Xu et al., 2016; Wang & Cai, 2016). Xu, et al. (Xu et al., 2016) focuses

on miRNA-transcription factor (TF)-mRNA regulatory networks, and they use their method to identify the miRNA-TF-mRNA regulatory network for clustering samples with different cancer subtypes and achieve the goal of cancer subtype classification. Noh, et al. (Noh et al., 2014) focuses on identifying a set of miRNA-mRNA connections that are changed in Alzheimer's disease. Wang, et al. (Wang & Cai, 2016) uses PageRank to rank miRNAs and mRNAs, separately, and to select the top ranked ones as biomarkers for ischemic stroke. Although these studies make use of PageRank, their focus is totally different from our approach.

Here, we propose a new PageRank-based method, called miR2Pathway, to rank disease risk of rewired miRNA-mediated biological pathways and we apply it to study HCC. For example, in a hypothetical case 1, a miRNA regulates several hub genes in a pathway in normal tissue, while it loses the regulatory connections in tumor tissue; in case 2, this miRNA regulates the same number of non-hub genes in this pathway in normal tissue, while it loses the regulatory connections in tumor tissue. In this scenario, our hypothesis is that this miRNA has a larger differential influence on the activity of the pathway in case 1 than case 2. This is also related to the basic idea of PageRank (Page et al., 1999) that the topology of a node is high in a network if this node has connections to other nodes with high topology. Using a PageRank-based approach, miR2Pathway focuses on quantifying the differential effects of miRNAs on the activity of a biological pathway when miRNA-mRNA connections are altered from normal to HCC. miR2Pathway provides a new insight to explore HCC mechanism. Thus, miR2Pathway is a novel method that can identify miRNA-dysregulated pathways in cancer and has several characteristics which are different from previous methods (Nam et al., 2009; Maragkakis et al., 2011; Hsu et al., 2011a; Lu et al., 2012; Ben-Hamo & Efroni, 2015; Godard & van Eyll, 2015; Preusse, Theis & Mueller, 2016; Backes et al., 2016; Han et al., 2016;

Backes et al., 2017): (1) It can identify the relationship between a set of miRNAs and a set of pathways. (2) It focuses on identifying changes in miRNA-mediated pathways between control and case, while the other methods focus on finding pathways enriched in genes targeted by miRNAs. (3) It focuses on identifying the topological changes in miRNA-mediated pathways between control and case, while the other methods do not assess or use topological changes. These characteristics, particularly (2) and (3), make it difficult to compare miR2Pathway with other methods because miR2Pathway addresses a different question from the other methods.

3.2. Materials and Methods

An overview of the miR2Pathway method is illustrated in **Figure 3.1**. Briefly, gene and microRNA expression profiles are used to construct connections between each miRNA and genes of each pathway for control and case, respectively. Subsequently, we obtain the corresponding differential network between control and case for each miRNA-Pathway pair. We can find the genes targeted by the miRNA in this differential network. Then, PageRank can be applied to measure the topological influence (PageRank scores) of the targeted genes in this differential network, which quantifies the topological influence of the genes that are differentially targeted by a miRNA on the activity of this pathway. Next, we can calculate the sum of PageRank scores of genes targeted by the miRNA in the differential network, which estimates the total differential influence of a miRNA on the activity of this pathway. Then, the same procedure is repeated for all miRNAs. We obtain a corresponding sum of PageRank scores for each miRNA. For a specific pathway, we then assess the total differential influence of all the miRNAs on this pathway through summing up all the sums corresponding to all the miRNAs. The total differential influence of all the miRNAs on this pathway reflects the degree of miRNA-

mediated dysregulation of this pathway. We do this for all the pathways. Finally, we rank all pathways by the degree of miRNA-mediated dysregulation scores.

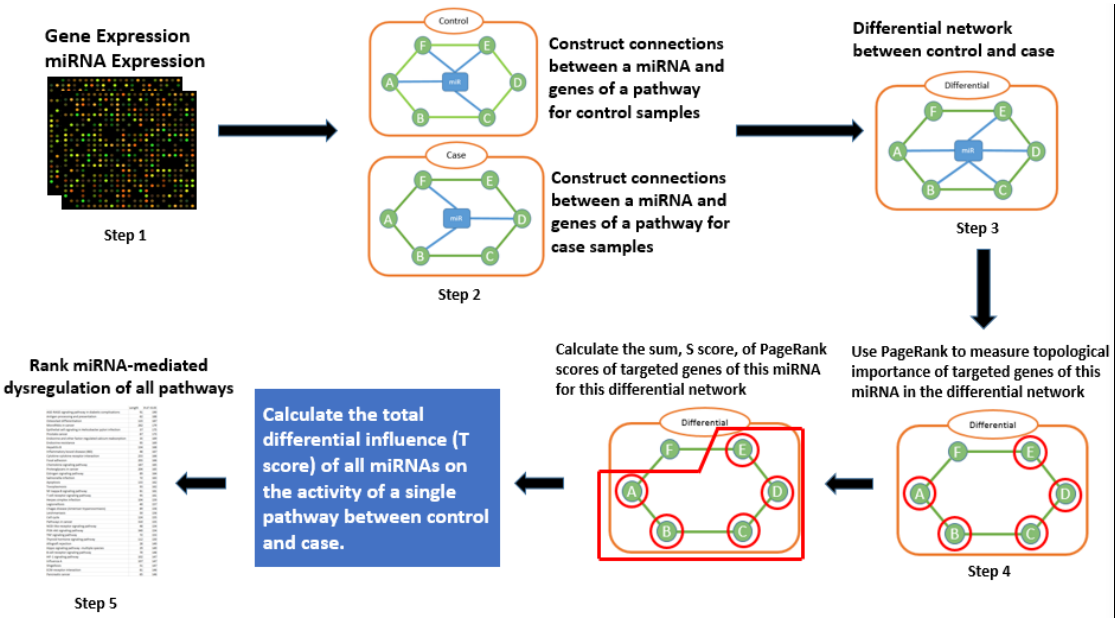


Figure 3.1. An overview of miR2Pathway.

3.2.1. Data

For illustration, we use RNA-seq data of matched miRNA and mRNA from The Cancer Genome Atlas (TCGA) hepatocellular carcinoma (HCC) study (https://cancergenome.nih.gov/). The dataset contains expression levels for 1,046 miRNAs and 20,531 mRNAs. We apply miR2Pathway to analyzing four datasets, containing 50 HCC samples and 50 tumor-adjacent normal samples, 34 hepatitis C-induced HCC samples and 34 tumor-adjacent normal samples, 22 hepatitis B-induced HCC samples and 22 tumor-adjacent normal samples and 50 alcohol-induced HCC samples and 50 tumor-adjacent normal samples. Tumor-adjacent normal samples in

following sections will be referred to as normal samples. Pathway information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa & Goto, 2000) is used.

3.2.2. Construct connections between each miRNA and genes of each pathway

3.2.2.1. Predicted and validated MiRNA targets

Five miRNA target site prediction programs (DIANA (Maragkakis et al., 2011), Targetscan (Friedman et al., 2009), PicTar (Lall et al., 2006), Miranda (Enright et al., 2003) and miRDB (Wong & Wang, 2015)) are employed to obtain putative miRNA target genes for all 1,046 miRNAs. These five programs are included in the "miRNAtap" R package. Therefore, in practice, we implement the miRNA target site prediction by using miRNAtap. We select potential target genes when they are identified by at least two of five programs. Three validated miRNA target site databases (miRecords (Xiao et al., 2009), miRTarBase (Hsu et al., 2011b) and TarBase (Vergoulis et al., 2012)) are used to obtain validated miRNA target genes for all 1,046 miRNAs. These three databases are included in the "multiMiR" R package and database. We select validated target genes when they are present in at least one of these three databases.

3.2.2.2. Statistical analysis of miRNAs and target genes

As we know, miRNA expression is negatively correlated with mRNA expression. To statistically identify the miRNA-mRNA connections in the regulatory network, we define a statistical connection between a miRNA and its target genes if the Pearson's correlation between them is less than a series of cutoffs (-0.4, -0.3, -0.2, -0.1) and the corresponding

p-value of the Pearson's correlation is < 0.05. We implement it through a built-in function called "cor.test()" in the statistical software package R (https://www.r-project.org/). This is done separately for case and control samples.

3.2.2.3. Identification of connections between miRNAs and mRNAs

Based on the results above, we use the intersection of the sets from **Step 2.1** and **Step 2.2** as the identified miRNA-mRNA connections, which are used for the following analysis.

Note, for Steps 2.1-2.3, we only determine the connections between miRNAs and target genes. For the construction of pathways, we directly obtain gene-gene connections within pathways from KEGG (Kanehisa & Goto, 2000). We pre-define the topology of pathways from KEGG and observe alterations of miRNA-mRNA connections in our study, thereby we can better quantify the degree by which miRNAs differentially influence the activity of each pathway between control and case.

3.2.2.4. Construct a miRNA-Pathway regulatory network

Next, we construct the miRNA-Pathway networks, i.e., construct each network consisting of a single miRNA and a single pathway. First, we obtain the gene list of a specified pathway from KEGG. Second, for each miRNA, we select the identified miRNA-mRNA connections specific for this pathway. Finally, we merge these identified miRNA-mRNA connections into the topology of this pathway derived from KEGG. The result is a miRNA-Pathway network. This is done separately for case and control samples.

3.2.3. Differential networks for miRNA-Pathway pairs

Based on the constructed miRNA-Pathway regulatory networks for control and case, separately, we can easily find the corresponding differential networks between control and case, see **Figure 3.2**.

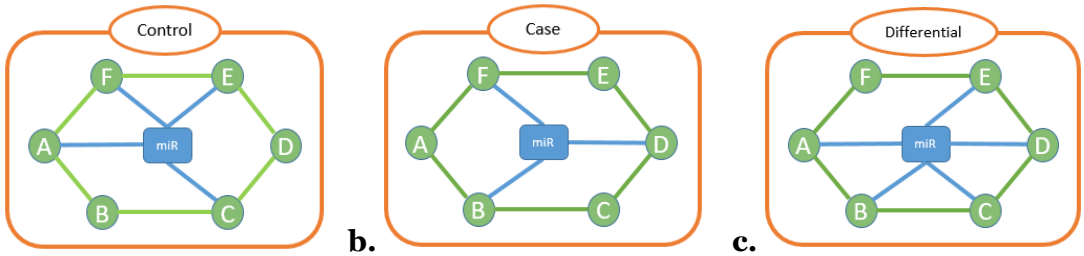


Figure 3.2. Construction of differential networks. (a) is a miRNA-Pathway regulatory network for control. (b) is the miRNA-Pathway regulatory network for case. (c) is the differential miRNA-Pathway regulatory network between control and case. A differential connection is constructed if it appears in either one of (a) and (b) while it does not appear in the other one. The green nodes represent genes and the blue node represents a miRNA.

Notably, all the networks in our study are based on correlation, thus, all of them are undirected graphs.

3.2.4. Measure the differential influence of miRNAs on the activity of pathways

3.2.4.1. Measure the differential influence of a single miRNA on a single pathway

To summarize, up to this point the algorithm finds the differential miRNA-Pathway network, which provides information about the differential influence of a miRNA on the activity of a pathway between control and case. The null hypothesis is that there is no difference in the miRNA-Pathway network between control and case, indicating that the

differential miRNA-Pathway network has an isolated miRNA. In other words, the single miRNA has no differential influence on the activity of this pathway. Conversely, if this miRNA has many differential connections in the miRNA-Pathway network between control and case, and, even more importantly, has differential connections with hub genes in this pathway, it suggests that this miRNA has a large differential influence on the activity of this pathway between control and case samples.

In this step, we measure the differential influence of a miRNA on the activity of this pathway between control and case using PageRank (Page et al., 1999), see Eq. (1).

$$S_{i,j} = PR(TG)_{i,j,1} + PR(TG)_{i,j,2} + \cdots + PR(TG)_{i,j,k} \quad (1)$$

In Eq. (1), $S_{i,j}$ is the sum of PageRank (PR) score of targeted genes (TG) of the miRNA i for the pathway j in the differential miRNA-Pathway network. The letter i denotes the index of a miRNA, the letter j denotes the index of a pathway and letter k denotes the number of targeted genes of the miRNA i in the corresponding differential miRNA-Pathway network. $S_{i,j}$ quantifies the differential influence of the miRNA i on the activity of the pathway j between control and case. Since PageRank considers the sum of PageRank scores of all the nodes in a network equal to 1, $S_{i,j}$ ranges from 0 to 1. When $S_{i,j} = 0$, it indicates that the miRNA i does not differentially regulate genes of the pathway j between control and case. When $S_{i,j} = 1$, it indicates that the miRNA i differentially regulates all genes of the pathway j between control and case.

3.2.4.2. Measure the total differential influence of a set of miRNAs on a single pathway

For this same pathway, we repeat the same procedure from Step 4.1 for all miRNAs of interest. We obtain different $S_{i,j}$ (i=1,2,…,M, where M is the number of miRNAs and j is the index of the pathway) for each miRNA. Then, we assess the total differential influence of all the miRNAs on this pathway through summing up all the $S_{i,j}$ scores, see Eq. (2).

$$T_j = S_{1,j} + S_{2,j} + \cdots + S_{M,j} \quad (2)$$

In Eq. (2), T_j is the sum of all the S scores corresponding to all the miRNAs for pathway j. The letter M represents the number of miRNAs, and the letter j represents the index of a pathway. The T score quantifies the total differential influence of all the miRNAs on the activity of a single pathway between control and case. If T_j is larger, it suggests that miRNAs differentially regulate a larger number of genes and/or differentially regulate hub genes in pathway j between control and case. Hence, the T score can reflect the degree of miRNAs dysregulating a single pathway.

3.2.5. Rank pathways based on disease risk of miRNA-mediated dysregulation of biological pathways

We repeat Step 4 to obtain a corresponding T score for each pathway. Finally, we rank all pathways by their T scores, which are measures of the degree of miRNA-mediated dysregulation.

3.2.6. Software tools

All the analysis is conducted using the R programming language. We use the following R Biocoductor packages: parallel for parallel computing, graphite for pathway databases,

igraph for PageRank function, graph for visualization, miRNAtap for miRNA target site prediction.

3.3. Results

First, we apply miR2Pathway to analyze HCC samples and normal samples. We use as example the interaction between miRNA-122 and the "MicroRNAs in cancer" pathway. MiRNA-122 is reported to be specific for liver cancer in several studies (Coulouarn et al., 2009; Li et al., 2012; Tsai et al., 2012; Bandiera et al., 2015), and "MicroRNAs in cancer" is a miRNA-related pathway. We are interested in seeing how miRNA-122 differentially influences the activity of this pathway. **Figure 3.3** shows the topological structure of the pathway "MicroRNAs in cancer".

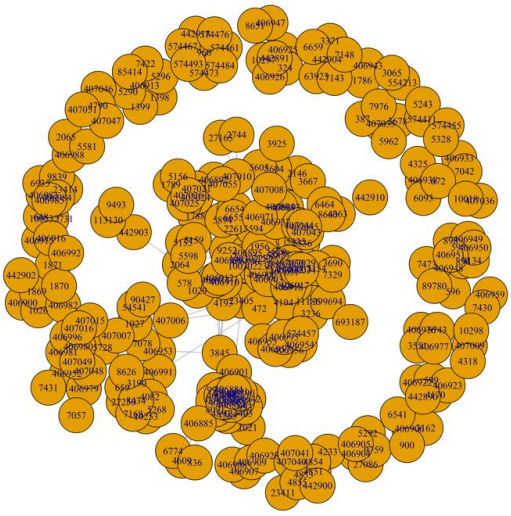


Figure 3.3. The pathway "MicroRNAs in cancer". We directly obtain the topological structure of this pathway from the KEGG database. There are 262 genes in this pathway, and there are 518 connections between genes in this pathway. The number inside each node is the corresponding gene ID. For gene names, see **Table 3.3**.

3.3.1. Construct connections between miRNA-122 and genes of the pathway "MicroRNAs in cancer" for normal and HCC

Based on the topological structure of this pathway, we need to know which gene(s) is/are targeted by miRNA-122 in this pathway for normal and cancer, separately. After we complete **Step 2.3**, we identify connections between miRNA-122 and genes of this pathway "MicroRNAs in cancer". In **Figure 3.4**, as an example, we only show the identified connections between miRNA-122 and genes based on predicted targets using a correlation cutoff of -0.4.

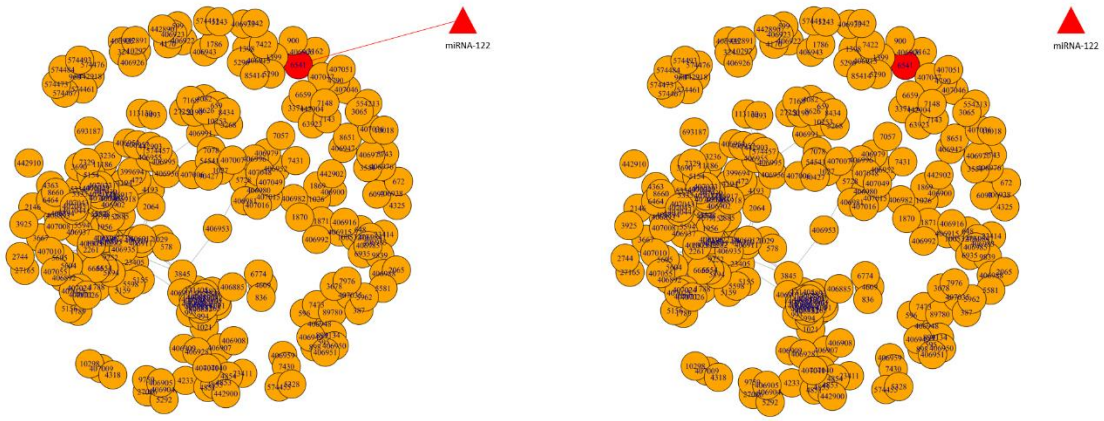


Figure 3.4. The interaction between miRNA-122 and the "MicroRNAs in Cancer" pathway for normal (a) and cancer (b), separately.

Figure 3.4 shows that miRNA-122 targets one gene (geneID is 6541 and gene symbol is CAT-1 or SLC7A1) in normal, while it does not target this gene in HCC. Very interestingly, several studies show that CAT-1/SLC7A1 is a well-known target gene of miRNA-122 (Chang et al., 2004; Yang & Kaye, 2009; Gatfield et al., 2009; Cirera et al., 2010; Li et al., 2012). CAT-1/SLC7A1 is an important protein for liver tissue. It is a carrier protein required in the regenerating liver for the transport of cationic amino acids

and polyamines in the late G1 phase, a process that is essential for liver cells to enter mitosis, and CAT-1/SLC7A1 is involved in amino acid metabolism (Chang et al., 2004). Besides, several studies show that miRNA-122's loss of function has been observed in liver cancer (Coulouarn et al., 2009; Hsu et al., 2012; Tsai et al., 2012; Thakral & Ghoshal, 2015). Thus, this result suggests that miRNA-122's loss of function probably leads to loss of a connection between miRNA-122 and CAT-1/SLC7A1 in the HCC samples. Therefore, this result is well consistent with the evidence from prior literature.

3.3.2. A differential network between normal and cancer

Based on the result illustrated in **Figure 3.4**, we can easily obtain the differential network between normal and cancer, whose topological structure is the same as **Figure 3.4(a)**.

3.3.3. Measure the differential influence of miRNAs on activities of this specified pathway between normal and cancer

Based on the differential network above, we assess the topological influence of genes dysregulated by miRNA-122 in this pathway through calculating PageRank scores of targeted genes in the differential network. Namely, the PageRank score of the gene CAT-1/SLC7A1 in the pathway is 0.002647686, which is also shown in **Table 3.3**. In **Table 3.3**, we also compute and include eigenvector centrality, which is another topological measure of the "MicroRNAs in cancer" pathway, which was discussed previously in literature (Mallik & Maulik, 2015). However, we only discuss and use PageRank scores for genes of each pathway in our study.

In this way, we measure the differential influence of all other miRNAs on the activity of the pathway "MicroRNAs in cancer", then we sum up the differential influence of all the miRNAs as a total one (i.e., T score). This total score, T = 0.139, is the degree of miRNA-mediated dysregulation of the pathway of "MicroRNAs in cancer".

3.3.4. Rank miRNA-mediated dysregulation of all pathways

Similarly, we obtain a corresponding total score T for each pathway. Then, we rank pathways using the T scores and show the top 50 pathways of miRNA-mediated dysregulation in **Table 3.1,** using a Pearson correlation cutoff value of -0.4 . As mentioned in **Materials and Methods**, we also tested several other correlation cutoffs: -0.3, -0.2, -0.1. The results are very similar to those from **Table 3.1** and are included in **Table 3.6**, **Table 3.7** and **Table 3.8**, respectively.

		Gene Count(L)	T score
1	FoxO signaling pathway	126	0.652
2	Circadian rhythm	31	0.418
3	Hedgehog signaling pathway	47	0.375
4	Notch signaling pathway	48	0.354
5	Hippo signaling pathway -multiple species	29	0.333
6	Dorso-ventral axis formation	13	0.315
7	Cytosolic DNA-sensing pathway	21	0.265
8	Thyroid cancer	28	0.257
9	Shigellosis	51	0.253
10	Inflammatory bowel disease (IBD)	48	0.242
11	RNA degradation	18	0.217
12	Toll-like receptor signaling pathway	104	0.207

13	AGE-RAGE signaling pathway in diabetic complications	91	0.203
14	Wnt signaling pathway	137	0.177
15	MAPK signaling pathway	252	0.175
16	Cocaine addiction	42	0.168
17	mTOR signaling pathway	144	0.159
18	Oocyte meiosis	120	0.155
19	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	0.154
20	Epithelial cell signaling in Helicobacter pylori infection	37	0.153
21	Insulin resistance	94	0.151
22	Pancreatic cancer	65	0.142
23	Steroid biosynthesis	20	0.142
24	MicroRNAs in cancer	262	0.139
25	HTLV-I infection	194	0.137
26	Progesterone-mediated oocyte maturation	89	0.129
27	Adipocytokine signaling pathway	63	0.128
28	Rap1 signaling pathway	208	0.127
29	Acute myeloid leukemia	57	0.115
30	Hepatitis C	97	0.114
31	Ether lipid metabolism	44	0.110
32	Hepatitis B	134	0.107
33	Leishmaniasis	50	0.101
34	Glyoxylate and dicarboxylate metabolism	26	0.097
35	Herpes simplex infection	104	0.095
36	Pantothenate and CoA biosynthesis	16	0.094
37	Huntington's disease	27	0.092
38	Vascular smooth muscle contraction	114	0.091

39	Antigen processing and presentation	62	0.091
40	Breast cancer	143	0.088
41	Calcium signaling pathway	179	0.085
42	Long-term potentiation	67	0.085
43	Chagas disease (American trypanosomiasis)	89	0.084
44	Estrogen signaling pathway	89	0.082
45	RIG-I-like receptor signaling pathway	48	0.082
46	p53 signaling pathway	68	0.080
47	Osteoclast differentiation	123	0.079
48	Toxoplasmosis	93	0.077
49	Gap junction	88	0.076
50	Vasopressin-regulated water reabsorption	22	0.075

Table 3.1. Top 50 pathways ranked by T score comparing normal with HCC samples based on the Pearson's correlation cutoff (-0.4).

Interestingly, we find in **Table 3.1** a large number of pathways associated with cancer in general and liver cancer specifically. First, the "microRNAs in cancer" pathway is listed in the top 50 pathways, which suggests that miRNA-mediated dysregulation is able to contribute to cancer (Garzon, Calin & Croce, 2009; Kwak, Iwasaki & Tomari, 2010; Jansson & Lund, 2012). FoxO signaling pathway is top-ranked and known to be involved in the regulation of the cell cycle, apoptosis, and metabolism. The second ranked pathway, circadian rhythm, is well known to be implicated in cancer (Fu & Kettner, 2013; Blakeman et al., 2016). Hedgehog signaling pathway is a major regulator of many fundamental processes in vertebrate embryonic development including stem cell maintenance, cell differentiation, tissue polarity and cell proliferation (Evangelista, Tian & de Sauvage, 2006; Rubin & de Sauvage, 2006; Gupta, Takebe & LoRusso, 2010; Gonnissen, Isebaert & Haustermans, 2015). Notch signaling pathway is one of the most commonly activated signaling pathways in cancer, and plays a key role in cell

proliferation, differentiation and survival (Capaccione & Pine, 2013; Guo et al., 2014; Yuan et al., 2015). Hippo signaling pathway is reported to be able to control organ size through regulating cell proliferation and apoptosis (Saucedo & Edgar, 2007; Pan, 2010). Wnt signaling pathway is a well-known pathway specific for liver cancer. The deregulation of the Wnt signaling pathway is early event in hepatocarcinogenesis and Wnt signaling pathway plays a critical role in liver development, regeneration, and promoting tumor formation in this organ (Takigawa & Brown, 2008; Waisberg & Saba, 2015; Wang et al., 2017). MAPK signaling pathway is involved in the regulation of survival, cellular growth, gene expression (Knight & Irving, 2014). Deregulation of MAPK signaling pathway can lead to uncontrolled or increased cell proliferation and resistance to apoptosis (Santarpia, Lippman & El-Naggar, 2012; Burotto et al., 2014). mTOR signaling pathway is a well-known cancer-associated pathway. Alterations of mTOR signaling pathway have significant effects on cancer progression. The major components of mTOR signaling pathway are critical effectors in cell signaling pathways commonly deregulated in cancers (Guertin & Sabatini, 2007; Villanueva et al., 2008; Pópulo, Lopes & Soares, 2012). Rap1 signaling is very important in basic cellular functions (e.g., formation, junctions and control of cell adhesions), cellular migration, and polarization (Bos, de Rooij & Reedquist, 2001). Rap1 plays key roles during cell invasion and metastasis in various cancers (Bos, de Rooij & Reedquist, 2001, p. 1; Zhang et al., 2017). p53 signaling pathway is a very important oncogenic pathway, and it can regulate apoptosis, the cell cycle and help prevent cancer. P53 protein, a major component of p53 signaling pathway, is most frequently altered in cancer (May & May, 1999).

Interestingly, Hepatitis C and Hepatitis B pathways are found within the top 50 pathways in **Table 3.1**. It is well known that Hepatitis C and Hepatitis B are major risk

factors for liver cancer (Beasley, 1988; Chen et al., 2008; Hoshida et al., 2014; Goossens & Hoshida, 2015).

Table 3.1 also includes several pathways that are immune- and inflammatory- related, such as Toll-like receptor signaling pathway, HTLV-I infection, Antigen processing and presentation and RIG-I-like receptor signaling pathway. It is well documented (Grivennikov, Greten & Karin, 2010; Greten, Duffy & Korangy, 2013; Bishayee, 2014; Sachdeva, Chawla & Arora, 2015) the fact that the immune system plays a key role in the development and progression of cancer, and inflammatory responses play critical roles at different stages of cancer development, including initiation, promotion, malignant conversion, invasion, and metastasis. In addition, inflammation affects immune surveillance and responses to therapy.

Table 3.1 also includes several other tumor-associated pathways, such as Thyroid cancer, Pancreatic cancer, Acute myeloid leukemia and Breast cancer.

3.3.5. miR2Pathway for cancer subtype analysis

We then apply miR2Pathway to all the pathways from KEGG comparing normal and hepatitis B-induced HCC samples. **Table 3.2** lists the top 50 pathways of miRNA-mediated dysregulation for this analysis. As before, here we only show and analyze the results for hepatitis B-induced HCC based on the Pearson's correlation cutoff of -0.4. The other results, based on different Pearson's correlation cutoff (-0.3, -0.2, -0.1), are shown in **Table 3.9**, **Table 3.10** and **Table 3.11**, respectively.

		Gene Count(L)	T score
1	Circadian rhythm	31	4.603

2	FoxO signaling pathway	126	4.547
3	Hedgehog signaling pathway	47	4.281
4	Dorso-ventral axis formation	13	3.619
5	GnRH signaling pathway	85	3.516
6	Toll-like receptor signaling pathway	104	2.513
7	Wnt signaling pathway	137	2.398
8	MAPK signaling pathway	252	2.388
9	Rap1 signaling pathway	208	2.379
10	Epithelial cell signaling in Helicobacter pylori infection	37	2.083
11	Shigellosis	51	2.058
12	Renal cell carcinoma	57	2.043
13	Notch signaling pathway	48	1.948
14	Estrogen signaling pathway	89	1.896
15	Calcium signaling pathway	179	1.796
16	Hippo signaling pathway -multiple species	29	1.733
17	Gap junction	88	1.694
18	Thyroid cancer	28	1.671
19	AGE-RAGE signaling pathway in diabetic complications	91	1.629
20	HTLV-I infection	194	1.612
21	**Bladder cancer**	29	1.394
22	Vascular smooth muscle contraction	114	1.343
23	**TNF signaling pathway**	72	1.282
24	**Amyotrophic lateral sclerosis (ALS)**	36	1.230
25	**D-Glutamine and D-glutamate metabolism**	4	1.181
26	Cocaine addiction	42	1.179
27	**Sulfur metabolism**	9	1.171
28	MicroRNAs in cancer	262	1.157

29	**Fc epsilon RI signaling pathway**	61	1.143
30	**Long-term depression**	59	1.125
31	Chagas disease (American trypanosomiasis)	89	1.108
32	**Salmonella infection**	72	1.098
33	RNA degradation	18	1.083
34	**Tight junction**	125	1.077
35	**Insulin signaling pathway**	139	1.010
36	Leishmaniasis	50	0.986
37	**Viral carcinogenesis**	6	0.959
38	mTOR signaling pathway	144	0.944
39	**Adherens junction**	71	0.924
40	Oocyte meiosis	120	0.916
41	**Bacterial invasion of epithelial cells**	57	0.913
42	**Tuberculosis**	173	0.912
43	Vasopressin-regulated water reabsorption	22	0.897
44	RIG-I-like receptor signaling pathway	48	0.893
45	**Inflammatory mediator regulation of TRP channels**	91	0.879
46	**Retrograde endocannabinoid signaling**	59	0.853
47	**Type II diabetes mellitus**	47	0.843
48	Progesterone-mediated oocyte maturation	89	0.826
49	**Thyroid hormone synthesis**	46	0.823
50	**Glutamatergic synapse**	89	0.815

Table 3.2. Top 50 pathways ranked by T score comparing normal with hepatitis B-induced HCC samples based on the Pearson's correlation cutoff (-0.4). The bold-font pathways are non-overlapped with top 50 pathways in **Table 3.1**.

We find that most of the top 50 pathways from **Table 3.2** overlap with the results from **Table 3.1**. The non-common pathways are marked by bold font. A number of these non-common pathways are theoretically specific for hepatitis B-induced HCC. As we know,

hepatitis B-induced HCC is mainly caused by hepatitis B virus, hence, those non-common pathways might be involved in inflammatory response and immune-system related. Interestingly, we find many of the non-common pathways are, indeed, inflammation- and immune-system related, such as TNF signaling pathway, Fc epsilon RI signaling pathway, Salmonella infection, Viral carcinogenesis, Bacterial invasion of epithelial cells, Inflammatory mediator regulation of TRP channels. The selective presence of inflammation-related pathways in **Table 3.2** suggests that miR2Pathway can identify pathways specific for subtypes of HCC.

We also identify in **Table 3.2** several other pathways that are involved in metabolism, such as D-Glutamine and D-glutamate metabolism, Sulfur metabolism, Insulin signaling pathway and Type II diabetes mellitus. Three of these four pathways are related to hepatitis B virus infection. Several studies show that glutamine synthesis and metabolism are potential markers of HCC patients infected by hepatitis B (Long et al., 2010; Li et al., 2015). Very interestingly, some studies show that hepatitis B virus infection can contribute to the impairment of insulin signaling (Kim, Kim & Cheong, 2010; Barthel et al., 2016). It is reported that hepatitis B virus infection rate is higher in type II diabetes mellitus patients compared with healthy controls, suggesting that hepatitis B virus infection is associated with type II diabetes (Demir et al., 2008).

We then apply miR2Pathway to analyze other subtypes of HCC, such as hepatitis C-induced and alcohol-induced HCC. The top 50 pathways are listed in **Table 3.4** and **Table 3.5**, respectively.

In **Table 3.4**, we include the top 50 pathways from the miR2Pathway analysis of hepatitis C-induced HCC based on the Pearson's correlation cutoff of -0.4. We again find that most of the pathways overlap with those from **Table 3.1**. The non-common pathways are theoretically specifically associated with hepatitis C virus (HCV) infection.

Interestingly, we discover several non-common pathways also listed in non-common pathways of **Table 3.2** (hepatitis B-induced HCC), such as D-Glutamine and D-glutamate metabolism, Bacterial invasion of epithelial cells and Insulin signaling pathway. HCV infection might increase glutamine use and dependence, and inhibiting glutamine metabolism attenuates HCV infection and the oxidative stress associated with HCV infection (Lévy et al., 2017). Some studies show that hepatitis C virus can induce insulin resistance (IR), thereby contributing to steatosis, progression of fibrosis and HCC (Sheikh et al., 2008; Bose, 2014). Additionally, Bacterial invasion of epithelial cells is an inflammation-related pathway, which is specifically associated with this subtype of liver cancer. The results for hepatitis C-induced HCC based on different Pearson's correlation cutoffs (-0.3, -0.2, -0.1) can be found in **Supplementary Table 3.12**, **Table 3.13** and **Table 3.14**, respectively. In **Table 3.5**, we include the top 50 pathways from the miR2Pathway analysis of alcohol-induced HCC based on the Pearson's correlation cutoff of -0.4. Again, we find that most of the pathways overlap with **Table 3.1**. Among the non-common pathways between **Table 3.5** and **Table 3.1**, we find many pathways associated with inflammation and immune system, such as Pathogenic Escherichia coli infection, TNF signaling pathway, B cell receptor signaling pathway and GnRH signaling pathway. Some studies have shown that the major mechanisms of alcohol-induced HCC include pathways of the immune system and inflammation, reviewed in (Sidharthan & Kottilil, 2014). Additionally, very interestingly, Glutamatergic synapse and GABAergic synapse are two of non-common pathways for this subtype of HCC. Several studies show that alcohol induces many neuroadaptative changes in the CNS involving both glutamatergic and GABAergic synaptic transmission (Dildy-Mayfield et al., 1996; Lovinger & Roberto, 2013). The results for alcohol-induced HCC based on different

Pearson's correlation cutoff (-0.3, -0.2, -0.1) are shown in **Table 3.15**, **Table 3.16** and **Table 3.17**, respectively.

3.4. Discussion

We propose a PageRank-based method, called miR2Pathway, to rank disease risk of miRNA-mediated biological pathways in HCC. miR2Pathway can help explore how much miRNAs differentially influence the activity of biological pathways between two classes of phenotypes. The basic idea of PageRank is that the topological importance of a node is high in a network if this node has connections with other nodes with high topological importance (Page et al., 1999). This can be well applied to miRNA-mediated biological pathways. For example, a miRNA has a larger differential influence on a pathway if it regulates more genes, particularly hub genes, in the differential network between cases and control. Based on this observation, we assess the differential influence of a miRNA on the activity of a pathway between normal and HCC through summing up the PageRank scores of targeted genes of this miRNA in the corresponding differential network, which we call S score. Then, we assess the differential influence of all other miRNAs on the activity of this same pathway between normal and HCC. We sum up the S scores for all miRNAs to obtain the T score, which measures the total differential influence of all the miRNAs on the activity of this pathway. In the same way, we calculate corresponding T scores for the total differential influence of all the miRNAs on the activity of each pathway. Finally, we rank all pathways by -the T scores, which are measures of the degree of miRNA-mediated dysregulation. The miR2Pathway method focuses on quantifying the differential effects of a set of miRNA on the activity of biological pathways when miRNA-mRNA connections are altered from normal to HCC.

Our use of PageRank to study the effect of miRNA-mRNA connections at the pathway level is novel. Previous uses of PageRank to study miRNA have been focused on clustering analysis for disease subtype classification in cancer (Xu et al., 2016) or identification of hub genes in Alzheimer's Disease (Noh et al., 2014) or ischemic stroke (Wang & Cai, 2016).

In our application of miR2Pathway to study HCC and its subtypes, we find that many highly ranked pathways are tumor-associated, such as FoxO signaling pathway, circadian rhythm, Wnt signaling pathway, MAPK signaling pathway, mTOR signaling pathway, p53 signaling pathway, etc., as well as HCC-specific pathways, such as Hepatitis B, Hepatitis C, etc. These results suggest that these important pathways are dysregulated by rewired miRNA-mRNA connections in cancer. Besides, many other pathways are associated with inflammation, immune system and metabolism, etc., which directly link to the occurrence and progression of HCC. In addition, we also find that the "MicroRNAs in cancer" pathway is listed in top 50 pathways, which is consistent with the fact that miRNAs dysregulate in cancer. Therefore, miR2Pathway can quantify and rank the dysregulation of these biological pathways.

Further, we apply miR2Pathway to analyze three subtypes of HCC: hepatitis B-induced HCC, hepatitis C-induced HCC and alcohol-induced HCC. By comparing each subtype of HCC with HCC, we check whether the non-common pathways from each subtype of HCC are indeed related to each specific subtype of HCC. For hepatitis B-induced HCC and hepatitis C-induced HCC, both hepatitis B and hepatitis C viruses are strongly linked to inflammation response and immune system. Related to this, among the non-common pathways for hepatitis B-induced HCC and hepatitis C-induced HCC, we indeed find several inflammation- and immune-related pathways, such as TNF signaling pathway, Fc epsilon RI signaling pathway, Salmonella infection, Viral carcinogenesis, Bacterial

invasion of epithelial cells, and Inflammatory mediator regulation of TRP channels. Notably, Viral carcinogenesis is strongly associated with virus-induced HCC. Similarly, for alcohol-induced HCC, we find several pathways related to immune system and inflammation, such as Pathogenic Escherichia coli infection, TNF signaling pathway, B cell receptor signaling pathway and GnRH signaling pathway. Previous studies have shown that these pathways of immune system and inflammation are related to the major mechanisms of alcohol-induced HCC (Sidharthan & Kottilil, 2014).

3.5. Future Directions

In this study, while we apply miR2pathway to HCC as a proof of concept, we intend to study other cancers and diseases to assess the generalizability of our method. We apply miR2Pathway to pre-defined biological pathways, from the well-curated KEGG pathway database. In addition to this application, miR2Pathway can also be applied to any set of genes of interest, such as functional gene networks. Moreover, miR2Pathway could also be used to assess differential influence of other regulatory factors, e.g., Transcriptional factors (TFs) or circular RNAs, on biological pathways. These could be interesting areas to further explore.

We can further validate the miR2Pathway method using simulated data. First, we select a specified pathway (e.g., “MicroRNAs in cancer”). Then, we can simulate miRNA-mRNA connections for two networks (e.g., network A and B). To start the simulation, we only consider one miRNA and only one miRNA-gene connection is rewired for both the network A and B, but it is one rewired miRNA-hub gene connection for the network A, while one rewired miRNA-non hub gene connection for the network B. Hence, the expected result is that T score of the network A is greater than that of the network B.

Then, we apply miR2Pathway to the simulated data to test if the expected result is obtained.

For the PageRank algorithm, we use the default value of damping factor 0.85 in this study. However, the value of damping factor might be varying in different cases. Hence, we might further use simulation to decide the value of damping factor in different cases. This is a very interesting and valuable direction to investigate.

We obtain ranks of pathways in this study, and we also can further investigate some common miRNAs that differentially influence on highly ranked pathway. Those common miRNAs also could be drive miRNAs to drive cancer. Extensively, we might further quantify each miRNA's differential influence on each pathway, then sum up total differential influence of each miRNA on all pathways, finally, we can rank all miRNAs. Those highly ranked miRNAs could be driver miRNAs to drive tumorigenesis. We might further investigate those key miRNAs and explore the biology behind those results.

3.6. Conclusion

In summary, miR2Pathway is a novel method that can be used to assess the total differential influence of all miRNAs on the activity of a single pathway between control and case. The total differential influence can reflect the degree of miRNA-mediated dysregulation of pathways, thereby assessing disease risk of miRNA-mediated biological pathways. We apply this method to study HCC and its subtypes and find that a number of highly ranked biological pathways are involved in cancer generally and HCC specifically. Also, we find many highly ranked pathways are related to inflammation, immune system and metabolism, etc., which are directly associated with pathogenesis of cancer. miR2Pathway is also able to identify pathways specific to HCC subtypes.

Therefore, miR2Pathway is a new method to explore dysregulated pathways by analyzing rewired miRNA-mRNA connections.

3.7. Availability

R software to carry out the miR2Pathway computation is available via http://www.dinulab.org/tools.

3.8. Supplementary Materials

Index	Gene ID	Gene Symbol	PageRank Score	Eigenvector Centrality
1	100302238	MIR103B1	0.003760521	0.041230011
2	100302282	MIR103B2	0.003760521	0.041230011
3	406881	MIRLET7A1	0.004941166	0.965748397
4	406882	MIRLET7A2	0.004941166	0.965748397
5	406883	MIRLET7A3	0.004941166	0.965748397
6	406884	MIRLET7B	0.004941166	0.965748397
7	406885	MIRLET7C	0.005280395	0.370605112
8	406886	MIRLET7D	0.004941166	0.965748397
9	406887	MIRLET7E	0.004941166	0.965748397
10	406888	MIRLET7F1	0.004941166	0.965748397
11	406889	MIRLET7F2	0.004941166	0.965748397
12	406890	MIRLET7G	0.004941166	0.965748397
13	406891	MIRLET7I	0.004941166	0.965748397
14	406892	MIR100	0.00116434	0.00674992
15	406893	MIR101-1	0.00300433	0.012099974
16	406894	MIR101-2	0.00300433	0.012099974
17	406895	MIR103A1	0.003760521	0.041230011
18	406896	MIR103A2	0.003760521	0.041230011
19	406900	MIR106B	0.003280572	2.03E-05
20	406901	MIR107	0.004914847	0.918751999
21	406902	MIR10A	0.006916144	0.041096829
22	406903	MIR10B	0.007923195	0.041464294
23	406904	MIR1-1	0.004754269	0.001911378
24	406905	MIR1-2	0.004754269	0.001911378

25	406906	MIR122	0.007324118	0
26	406907	MIR124-1	0.001082422	0.09057311
27	406908	MIR124-2	0.001082422	0.09057311
28	406909	MIR124-3	0.001082422	0.09057311
29	406910	MIR125A	0.004168031	0.118425226
30	406911	MIR125B1	0.00491907	0.120562116
31	406912	MIR125B2	0.00491907	0.120562116
32	406913	MIR126	0.012585104	0
33	406915	MIR128-1	0.003364869	1.06E-05
34	406916	MIR128-2	0.003364869	1.06E-05
35	406917	MIR129-1	0.00471112	0.032238736
36	406918	MIR129-2	0.00471112	0.032238736
37	406922	MIR133A1	0.00323224	8.93E-18
38	406923	MIR133A2	0.00323224	3.06E-17
39	406925	MIR135A1	0.00323224	4.15E-17
40	406926	MIR135A2	0.00323224	1.86E-17
41	406928	MIR137	0.001082422	0.09057311
42	406933	MIR141	0.003816794	0
43	406935	MIR143	0.009949437	0.139702723
44	406937	MIR145	0.004220106	0.034768161
45	406938	MIR146A	0.007324118	0
46	406943	MIR152	0.003816794	0
47	406947	MIR155	0.003816794	0
48	406948	MIR15A	0.008110503	1.27E-16
49	406949	MIR15B	0.00413922	1.76E-16
50	406950	MIR16-1	0.00413922	1.61E-16
51	406951	MIR16-2	0.00413922	1.93E-16
52	406952	MIR17	0.003510622	0.001091129
53	406953	MIR18A	0.002202202	0.095220707
54	406954	MIR181A2	0.001257143	0.005306336
55	406955	MIR181B1	0.001257143	0.005306336
56	406956	MIR181B2	0.001257143	0.005306336
57	406957	MIR181C	0.001257143	0.005306336
58	406959	MIR183	0.003816794	4.27E-19
59	406971	MIR195	0.00846517	0.03836491
60	406976	MIR199A1	0.003816794	7.95E-18
61	406977	MIR199A2	0.003816794	0
62	406979	MIR19A	0.003510622	0.001091129
63	406980	MIR19B1	0.001755769	0.001081459
64	406981	MIR19B2	0.001755769	0.001081459
65	406982	MIR20A	0.007049021	0.001121961
66	406983	MIR200A	0.005013964	4.34E-07

67	406984	MIR200B	0.005013964	4.34E-07
68	406985	MIR200C	0.005013964	4.34E-07
69	406988	MIR205	0.004825867	1.19E-08
70	406991	MIR21	0.017313782	0.007094906
71	406992	MIR210	0.001652493	1.02E-05
72	406995	MIR181A1	0.001257143	0.005306336
73	406996	MIR214	0.001755769	0.001081459
74	407006	MIR221	0.008132586	0.007539487
75	407007	MIR222	0.004013834	0.001301028
76	407008	MIR223	0.002528401	0.006900102
77	407009	MIR224	0.005570456	6.98E-18
78	407010	MIR23A	0.004469908	0.008035219
79	407015	MIR26A1	0.001755769	0.001081459
80	407016	MIR26A2	0.001755769	0.001081459
81	407021	MIR29A	0.003268751	0.008938579
82	407024	MIR29B1	0.003268751	0.008938579
83	407025	MIR29B2	0.003268751	0.008938579
84	407026	MIR29C	0.003268751	0.008938579
85	407029	MIR30A	0.0060909	0.045466633
86	407030	MIR30B	0.004886337	0.043049027
87	407031	MIR30C1	0.004886337	0.043049027
88	407032	MIR30C2	0.004886337	0.043049027
89	407033	MIR30D	0.0060909	0.045466633
90	407034	MIR30E	0.0060909	0.045466633
91	407035	MIR31	0.00907778	0
92	407036	MIR32	0.003816794	0
93	407040	MIR34A	0.006423954	0.100639267
94	407041	MIR34B	0.005274269	0.099747382
95	407043	MIR7-1	0.004623953	0.030119613
96	407044	MIR7-2	0.004623953	0.030119613
97	407045	MIR7-3	0.004623953	0.030119613
98	407046	MIR9-1	0.002647686	6.42E-18
99	407047	MIR9-2	0.002647686	1.30E-17
100	407048	MIR92A1	0.001755769	0.001081459
101	407049	MIR92A2	0.001755769	0.001081459
102	407051	MIR9-3	0.002647686	1.68E-17
103	407055	MIR99A	0.00116434	0.00674992
104	442890	MIR133B	0.00323224	3.77E-18
105	442891	MIR135B	0.00323224	2.68E-17
106	442900	MIR326	0.003888798	0.007364528
107	442902	MIR330	0.001935407	1.02E-05
108	442903	MIR331	0.004467768	0.003273428

109	442904	MIR335	0.010831442	0
110	442910	MIR345	0.001284055	0.000807938
111	442918	MIR373	0.002313655	4.83E-17
112	554213	MIR449A	0.003816794	0
113	574411	MIR451A	0.003816794	0
114	574455	MIR193B	0.003816794	0
115	574457	MIR181D	0.001257143	0.005306336
116	574461	MIR520E	0.002313655	6.19E-17
117	574467	MIR520A	0.002313655	5.67E-17
118	574473	MIR520B	0.002313655	4.58E-17
119	574476	MIR520C	0.002313655	5.93E-17
120	574484	MIR520G	0.002313655	5.54E-17
121	574493	MIR520H	0.002313655	5.81E-17
122	693187	MIR602	0.001415719	0.000738215
123	2261	FGFR3	0.013925189	0.071701442
124	23405	DICER1	0.003083264	0.102016071
125	5578	PRKCA	0.011745384	0.08290846
126	5579	PRKCB	0.011745384	0.08290846
127	5582	PRKCG	0.011745384	0.08290846
128	9252	RPS6KA5	0.002703481	0.015525449
129	1021	CDK6	0.009598166	0.962118469
130	10642	IGF2BP1	0.004390693	0.909148282
131	131405	TRIM71	0.004390693	0.909148282
132	3265	HRAS	0.005031883	0.944036769
133	3845	KRAS	0.008403462	1
134	4893	NRAS	0.005031883	0.944036769
135	8091	HMGA2	0.005031883	0.944036769
136	993	CDC25A	0.004390693	0.909148282
137	994	CDC25B	0.004390693	0.909148282
138	995	CDC25C	0.004390693	0.909148282
139	4609	MYC	0.00121371	0.034888487
140	6774	STAT3	0.00121371	0.034888487
141	836	CASP3	0.00121371	0.034888487
142	2146	EZH2	0.001849359	0.002278165
143	5335	PLCG1	0.004960059	0.027276558
144	5336	PLCG2	0.004960059	0.027276558
145	1026	CDKN1A	0.003165096	0.000107535
146	1869	E2F1	0.004810192	0.000108497
147	11186	RASSF1	0.002975999	0.007841737
148	2885	GRB2	0.003570894	0.017453747
149	4193	MDM2	0.002760166	0.008482004
150	4194	MDM4	0.002916767	0.013842103

151	472	ATM	0.015303357	0.056366883
152	5594	MAPK1	0.00337662	0.012140315
153	3236	HOXD10	0.001184766	0.003903417
154	27086	FOXP1	0.002593084	0.000359871
155	4233	MET	0.004120323	0.01922412
156	5292	PIM1	0.002593084	0.000359871
157	9759	HDAC4	0.002593084	0.000359871
158	3162	HMOX1	0.002647686	1.37E-17
159	6541	SLC7A1	0.002647686	1.74E-17
160	900	CCNG1	0.002647686	1.00E-17
161	1029	CDKN2A	0.002357622	0.033847716
162	1956	EGFR	0.00463991	0.045627078
163	2064	ERBB2	0.003623489	0.034155874
164	7157	TP53	0.007193432	0.072647189
165	578	BAK1	0.00176715	0.022699254
166	1398	CRK	0.002355409	5.34E-17
167	1399	CRKL	0.002355409	3.51E-17
168	5290	PIK3CA	0.002355409	6.05E-17
169	5296	PIK3R2	0.002355409	4.82E-17
170	7422	VEGFA	0.002355409	5.79E-17
171	85414	SLC45A3	0.002355409	4.76E-17
172	100532731	COMMD3-BMI1	0.0050364	2.12E-06
173	1871	E2F3	0.005082231	0.000108583
174	648	BMI1	0.0050364	2.12E-06
175	399694	SHC4	0.001716648	0.006069861
176	4170	MCL1	0.004693625	0
177	599	BCL2L2	0.004693625	1.61E-17
178	10297	APC2	0.004693625	2.37E-17
179	324	APC	0.004693625	0
180	7042	TGFB2	0.003816794	0
181	1788	DNMT3A	0.004001498	0.016517394
182	5155	PDGFB	0.001223059	0.013151509
183	5159	PDGFRB	0.001223059	0.013151509
184	5598	MAPK7	0.001223059	0.013151509
185	5894	RAF1	0.001877186	0.016763153
186	6654	SOS1	0.001877186	0.016763153
187	6655	SOS2	0.001877186	0.016763153
188	3667	IRS1	0.001170367	0.003273049
189	4325	MMP16	0.002647686	8.78E-18
190	6093	ROCK1	0.002647686	1.70E-17
191	672	BRCA1	0.002647686	1.28E-17

192	1786	DNMT1	0.003816794	0
193	8651	SOCS1	0.003816794	0
194	595	CCND1	0.004196118	1.16E-16
195	596	BCL2	0.001557366	6.87E-17
196	7473	WNT3	0.001557366	6.90E-17
197	894	CCND2	0.004196118	1.97E-16
198	89780	WNT3A	0.001557366	7.51E-17
199	898	CCNE1	0.004196118	1.64E-16
200	9134	CCNE2	0.004196118	1.56E-16
201	5728	PTEN	0.019488814	0.011487865
202	7431	VIM	0.002064533	0.000102718
203	7430	EZR	0.003816794	0
204	5604	MAP2K1	0.001226646	0.003611644
205	5605	MAP2K2	0.001226646	0.003611644
206	3551	IKBKB	0.003816794	1.36E-17
207	5743	PTGS2	0.003816794	7.34E-18
208	7057	THBS1	0.002064533	0.000102718
209	1870	E2F2	0.001770853	0.000105621
210	23414	ZFPM2	0.003129641	1.23E-07
211	6935	ZEB1	0.003129641	1.23E-07
212	9839	ZEB2	0.00449697	1.24E-07
213	2065	ERBB3	0.001939848	1.12E-09
214	5581	PRKCE	0.001939848	1.12E-09
215	10253	SPRY2	0.001798912	0.000667909
216	27250	PDCD4	0.001798912	0.000667909
217	3190	HNRNPK	0.001798912	0.000667909
218	4082	MARCKS	0.001798912	0.000667909
219	5268	SERPINB5	0.001798912	0.000667909
220	659	BMPR2	0.001798912	0.000667909
221	7078	TIMP3	0.003923693	0.001500149
222	7168	TPM1	0.001798912	0.000667909
223	8434	RECK	0.001798912	0.000667909
224	8626	TP63	0.001798912	0.000667909
225	1027	CDKN1B	0.0026973	0.000832239
226	54541	DDIT4	0.001560047	0.000709762
227	90427	BMF	0.001560047	0.000709762
228	3925	STMN1	0.00164709	0.00064957
229	10298	PAK4	0.002939963	1.98E-17
230	4318	MMP9	0.002939963	1.40E-17
231	27165	GLS2	0.001522375	0.00075643
232	2744	GLS	0.001522375	0.00075643
233	1789	DNMT3B	0.003350958	0.003365884

234	5156	PDGFRA	0.003350958	0.003365884
235	3690	ITGB3	0.002125698	0.012840584
236	5154	PDGFA	0.003683218	0.024998393
237	7329	UBE2I	0.002125698	0.012840584
238	3678	ITGA5	0.002501547	3.23E-17
239	387	RHOA	0.002501547	3.96E-17
240	5962	RDX	0.002501547	3.55E-17
241	7976	FZD3	0.002501547	2.79E-17
242	10018	BCL2L11	0.003816794	0
243	23411	SIRT1	0.001352571	0.009474105
244	4851	NOTCH1	0.002926128	0.01955754
245	4853	NOTCH2	0.002926128	0.01955754
246	4854	NOTCH3	0.002926128	0.01955754
247	4855	NOTCH4	0.002926128	0.01955754
248	4363	ABCC1	0.003348406	0.008582372
249	6464	SHC1	0.002256959	0.008506313
250	8660	IRS2	0.002256959	0.008506313
251	4790	NFKB1	0.007324118	0
252	113130	CDCA5	0.001838387	0.000308158
253	9493	KIF23	0.001838387	0.000308158
254	3371	TNC	0.002413864	6.88E-17
255	63923	TNN	0.002413864	5.51E-17
256	6659	SOX4	0.002413864	5.43E-17
257	7143	TNR	0.002413864	5.69E-17
258	7148	TNXB	0.002413864	3.99E-17
259	960	CD44	0.014338766	0
260	3065	HDAC1	0.003816794	0
261	5243	ABCB1	0.003816794	0
262	5328	PLAU	0.003816794	0

Table 3.3. The information of the 262 genes in the "MicroRNAs in cancer" pathway. We consolidate the information as shown in this table: (1) Index (1-262) (2) Gene ID (3) Gene Symbol (4) The corresponding PageRank scores of each gene (5) The corresponding Eigenvector centrality of genes in this pathway, which is computed by a built-in function in the igraph R package called eigen_centrality().

		Gene Count(L)	T score
1	FoxO signaling pathway	126	2.519
2	Rap1 signaling pathway	208	1.523
3	Hedgehog signaling pathway	47	1.248

4	MicroRNAs in cancer	262	0.797
5	Vascular smooth muscle contraction	114	0.790
6	Hippo signaling pathway -multiple species	29	0.776
7	Circadian rhythm	31	0.753
8	**Renal cell carcinoma**	57	0.725
9	Wnt signaling pathway	137	0.707
10	mTOR signaling pathway	144	0.607
11	Progesterone-mediated oocyte maturation	89	0.604
12	Shigellosis	51	0.602
13	**D-Glutamine and D-glutamate metabolism**	4	0.590
14	Inflammatory bowel disease (IBD)	48	0.582
15	MAPK signaling pathway	252	0.574
16	AGE-RAGE signaling pathway in diabetic complications	91	0.527
17	Insulin resistance	94	0.511
18	Adipocytokine signaling pathway	63	0.510
19	RNA degradation	18	0.488
20	Estrogen signaling pathway	89	0.451
21	Pancreatic cancer	65	0.439
22	**Bacterial invasion of epithelial cells**	57	0.438
23	**Nitrogen metabolism**	4	0.435
24	Steroid biosynthesis	20	0.426
25	Cocaine addiction	42	0.421
26	HTLV-I infection	194	0.415
27	Notch signaling pathway	48	0.413
28	**alpha-Linolenic acid metabolism**	25	0.408
29	**Bladder cancer**	29	0.401
30	Cytosolic DNA-sensing pathway	21	0.398
31	Oocyte meiosis	120	0.392

32	Thyroid cancer	28	0.386
33	**Insulin signaling pathway**	139	0.377
34	Vasopressin-regulated water reabsorption	22	0.374
35	Toll-like receptor signaling pathway	104	0.355
36	**Leukocyte transendothelial migration**	85	0.348
37	**Parkinson's disease**	29	0.345
38	Epithelial cell signaling in Helicobacter pylori infection	37	0.344
39	Calcium signaling pathway	179	0.341
40	**Amyotrophic lateral sclerosis (ALS)**	36	0.328
41	Breast cancer	143	0.315
42	Long-term potentiation	67	0.314
43	**Signaling pathways regulating pluripotency of stem cells**	112	0.297
44	**Axon guidance**	167	0.294
45	**Long-term depression**	59	0.290
46	GnRH signaling pathway	85	0.288
47	**Proximal tubule bicarbonate reclamation**	7	0.286
48	Hepatitis C	97	0.282
49	**Cell cycle**	124	0.279
50	**Tuberculosis**	173	0.276

Table 3.4. Top 50 pathways ranked by T score comparing normal with hepatitis C-induced HCC samples based on the Pearson's correlation cutoff (-0.4). The bold-font pathways are non-overlapped with top 50 pathways in **Table 3.1**.

		Gene Count(L)	T score
1	FoxO signaling pathway	126	0.723
2	**alpha-Linolenic acid metabolism**	25	0.408
3	Notch signaling pathway	48	0.354

4	Hippo signaling pathway -multiple species	29	0.333
5	Hedgehog signaling pathway	47	0.322
6	Cytosolic DNA-sensing pathway	21	0.265
7	Circadian rhythm	31	0.251
8	Inflammatory bowel disease (IBD)	48	0.242
9	Toll-like receptor signaling pathway	104	0.237
10	Shigellosis	51	0.222
11	RNA degradation	18	0.217
12	Epithelial cell signaling in Helicobacter pylori infection	37	0.191
13	MAPK signaling pathway	252	0.167
14	AGE-RAGE signaling pathway in diabetic complications	91	0.162
15	mTOR signaling pathway	144	0.159
16	Progesterone-mediated oocyte maturation	89	0.151
17	HTLV-I infection	194	0.150
18	Oocyte meiosis	120	0.143
19	Insulin resistance	94	0.132
20	Adipocytokine signaling pathway	63	0.128
21	Rap1 signaling pathway	208	0.127
22	Wnt signaling pathway	137	0.126
23	Chagas disease (American trypanosomiasis)	89	0.117
24	Hepatitis C	97	0.114
25	MicroRNAs in cancer	262	0.111
26	**Long-term depression**	59	0.109
27	Pancreatic cancer	65	0.107
28	Leishmaniasis	50	0.101
29	Acute myeloid leukemia	57	0.100
30	Glyoxylate and dicarboxylate metabolism	26	0.097
31	Pantothenate and CoA biosynthesis	16	0.094

32	**Morphine addiction**	54	0.091
33	**Retrograde endocannabinoid signaling**	59	0.091
34	Hepatitis B	134	0.087
35	Toxoplasmosis	93	0.086
36	Herpes simplex infection	104	0.086
37	Long-term potentiation	67	0.085
38	Estrogen signaling pathway	89	0.082
39	RIG-I-like receptor signaling pathway	48	0.082
40	**Glutamatergic synapse**	89	0.080
41	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	0.077
42	**GABAergic synapse**	66	0.075
43	Osteoclast differentiation	123	0.069
44	**Pathogenic Escherichia coli infection**	40	0.068
45	**Bladder cancer**	29	0.067
46	Thyroid cancer	28	0.064
47	**TNF signaling pathway**	72	0.064
48	**B cell receptor signaling pathway**	70	0.062
49	**GnRH signaling pathway**	85	0.062
50	Vascular smooth muscle contraction	114	0.061

Table 3.5. Top 50 pathways ranked by T score comparing normal with alcohol-induced HCC samples based on the Pearson's correlation cutoff (-0.4). The bold-font pathways are non-overlapped with top 50 pathways in **Table 3.1**.

		Gene Count(L)	T score
1	FoxO signaling pathway	126	4.769631872
2	Cocaine addiction	42	3.790618613
3	Circadian rhythm	31	3.59848471
4	Hedgehog signaling pathway	47	3.247139618
5	Dorso-ventral axis formation	13	2.832281658
6	Notch signaling pathway	48	2.715229884
7	Rap1 signaling pathway	208	2.696510591

8	Hippo signaling pathway -multiple species	29	2.494673335
9	RNA degradation	18	2.383674555
10	Estrogen signaling pathway	89	2.166927109
11	Shigellosis	51	1.868168859
12	Calcium signaling pathway	179	1.84637558
13	Wnt signaling pathway	137	1.792202489
14	Vascular smooth muscle contraction	114	1.738601876
15	Vasopressin-regulated water reabsorption	22	1.719795843
16	MAPK signaling pathway	252	1.5537004
17	HTLV-I infection	194	1.531839998
18	Insulin resistance	94	1.419444038
19	Glutamatergic synapse	89	1.405421018
20	mTOR signaling pathway	144	1.403923184
21	Bacterial invasion of epithelial cells	57	1.388448828
22	Renal cell carcinoma	57	1.384068279
23	Thyroid hormone synthesis	46	1.283676831
24	Thyroid cancer	28	1.221223406
25	MicroRNAs in cancer	262	1.206816051
26	Long-term depression	59	1.19784699
27	Huntington's disease	27	1.149259079
28	GnRH signaling pathway	85	1.13073527
29	Gap junction	88	1.123257389
30	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	1.078378378
31	Progesterone-mediated oocyte maturation	89	1.055398393
32	AGE-RAGE signaling pathway in diabetic complications	91	1.054079381
33	Retrograde endocannabinoid signaling	59	1.034413371
34	Fatty acid biosynthesis	13	1.022351755
35	Pancreatic secretion	30	0.987496865
36	Oocyte meiosis	120	0.986979296
37	Amyotrophic lateral sclerosis (ALS)	36	0.983800494
38	Aldosterone synthesis and secretion	65	0.976533251
39	Maturity onset diabetes of the young	24	0.963574619
40	Viral carcinogenesis	6	0.959459459
41	Signaling pathways regulating pluripotency of stem cells	112	0.954367248
42	Long-term potentiation	67	0.942978837
43	Toll-like receptor signaling pathway	104	0.933864231
44	Melanogenesis	101	0.924727007
45	Insulin secretion	54	0.900427584
46	Morphine addiction	54	0.889599249

47	Epithelial cell signaling in Helicobacter pylori infection	37	0.880104219
48	Cytosolic DNA-sensing pathway	21	0.862616564
49	Inflammatory bowel disease (IBD)	48	0.858072477
50	Alzheimer's disease	48	0.843041679

Table 3.6. Top 50 pathways ranked by T score comparing normal with HCC samples based on the Pearson's correlation cutoff (-0.3).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	7.511864547
2	Cocaine addiction	42	6.038087082
3	Dorso-ventral axis formation	13	5.664563315
4	Circadian rhythm	31	5.439569911
5	Hedgehog signaling pathway	47	4.77862805
6	Rap1 signaling pathway	208	4.307901326
7	Hippo signaling pathway -multiple species	29	3.82516578
8	Estrogen signaling pathway	89	3.581660156
9	Notch signaling pathway	48	3.423550724
10	RNA degradation	18	3.19629088
11	Calcium signaling pathway	179	3.112523145
12	Wnt signaling pathway	137	2.85237861
13	Vasopressin-regulated water reabsorption	22	2.727502482
14	Shigellosis	51	2.659765832
15	Vascular smooth muscle contraction	114	2.65643198
16	Renal cell carcinoma	57	2.636320531
17	Thyroid cancer	28	2.442446812
18	MAPK signaling pathway	252	2.386014506
19	HTLV-I infection	194	2.26229856
20	Glutamatergic synapse	89	2.203955688
21	Thyroid hormone synthesis	46	2.166772169
22	mTOR signaling pathway	144	2.153472537
23	AGE-RAGE signaling pathway in diabetic complications	91	2.108158762
24	Insulin resistance	94	2.055957205
25	Long-term depression	59	2.032710043
26	GnRH signaling pathway	85	2.014764663
27	Huntington's disease	27	1.838814526
28	Gap junction	88	1.82767304
29	Bacterial invasion of epithelial cells	57	1.826906352
30	Pancreatic secretion	30	1.777494357
31	MicroRNAs in cancer	262	1.724442638
32	Amyotrophic lateral sclerosis (ALS)	36	1.680659178

33	Maturity onset diabetes of the young	24	1.651842204
34	Aldosterone synthesis and secretion	65	1.647899861
35	Progesterone-mediated oocyte maturation	89	1.622806724
36	Retrograde endocannabinoid signaling	59	1.578841462
37	Tight junction	125	1.577960616
38	Insulin secretion	54	1.560946654
39	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	1.540540541
40	Melanogenesis	101	1.491495172
41	mRNA surveillance pathway	70	1.486293436
42	Signaling pathways regulating pluripotency of stem cells	112	1.476804152
43	Bladder cancer	29	1.470584223
44	Long-term potentiation	67	1.450736673
45	Insulin signaling pathway	139	1.44934286
46	Salivary secretion	48	1.401201983
47	Morphine addiction	54	1.391424466
48	Oocyte meiosis	120	1.379392752
49	Toll-like receptor signaling pathway	104	1.377522425
50	Transcriptional misregulation in cancer	19	1.315789474

Table 3.7. Top 50 pathways ranked by T score comparing normal with HCC samples based on the Pearson's correlation cutoff (-0.2).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	7.511864547
2	Cocaine addiction	42	6.038087082
3	Dorso-ventral axis formation	13	5.664563315
4	Circadian rhythm	31	5.439569911
5	Hedgehog signaling pathway	47	4.77862805
6	Rap1 signaling pathway	208	4.307901326
7	Hippo signaling pathway -multiple species	29	3.82516578
8	Estrogen signaling pathway	89	3.581660156
9	Notch signaling pathway	48	3.423550724
10	RNA degradation	18	3.19629088
11	Calcium signaling pathway	179	3.112523145
12	Wnt signaling pathway	137	2.85237861
13	Vasopressin-regulated water reabsorption	22	2.727502482
14	Shigellosis	51	2.659765832
15	Vascular smooth muscle contraction	114	2.65643198
16	Renal cell carcinoma	57	2.636320531
17	Thyroid cancer	28	2.442446812
18	MAPK signaling pathway	252	2.386014506

19	HTLV-I infection	194	2.26229856
20	Glutamatergic synapse	89	2.203955688
21	Thyroid hormone synthesis	46	2.166772169
22	mTOR signaling pathway	144	2.153472537
23	AGE-RAGE signaling pathway in diabetic complications	91	2.108158762
24	Insulin resistance	94	2.055957205
25	Long-term depression	59	2.032710043
26	GnRH signaling pathway	85	2.014764663
27	Huntington's disease	27	1.838814526
28	Gap junction	88	1.82767304
29	Bacterial invasion of epithelial cells	57	1.826906352
30	Pancreatic secretion	30	1.777494357
31	MicroRNAs in cancer	262	1.724442638
32	Amyotrophic lateral sclerosis (ALS)	36	1.680659178
33	Maturity onset diabetes of the young	24	1.651842204
34	Aldosterone synthesis and secretion	65	1.647899861
35	Progesterone-mediated oocyte maturation	89	1.622806724
36	Retrograde endocannabinoid signaling	59	1.578841462
37	Tight junction	125	1.577960616
38	Insulin secretion	54	1.560946654
39	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	1.540540541
40	Melanogenesis	101	1.491495172
41	mRNA surveillance pathway	70	1.486293436
42	Signaling pathways regulating pluripotency of stem cells	112	1.476804152
43	Bladder cancer	29	1.470584223
44	Long-term potentiation	67	1.450736673
45	Insulin signaling pathway	139	1.44934286
46	Salivary secretion	48	1.401201983
47	Morphine addiction	54	1.391424466
48	Oocyte meiosis	120	1.379392752
49	Toll-like receptor signaling pathway	104	1.377522425
50	Transcriptional misregulation in cancer	19	1.315789474

Table 3.8. Top 50 pathways ranked by T score comparing normal with HCC samples based on the Pearson's correlation cutoff (-0.1).

		Gene Count(L)	T score
1	Circadian rhythm	31	4.602713001
2	FoxO signaling pathway	126	4.546560009
3	Hedgehog signaling pathway	47	4.280954331

4	Dorso-ventral axis formation	13	3.619026562
5	GnRH signaling pathway	85	3.515558749
6	Toll-like receptor signaling pathway	104	2.51299444
7	Wnt signaling pathway	137	2.398017415
8	MAPK signaling pathway	252	2.387515249
9	Rap1 signaling pathway	208	2.379274051
10	Epithelial cell signaling in Helicobacter pylori infection	37	2.082740987
11	Shigellosis	51	2.058152132
12	Renal cell carcinoma	57	2.043148412
13	Notch signaling pathway	48	1.947882308
14	Estrogen signaling pathway	89	1.895632946
15	Calcium signaling pathway	179	1.795913644
16	Hippo signaling pathway -multiple species	29	1.732928622
17	Gap junction	88	1.694405214
18	Thyroid cancer	28	1.671147819
19	AGE-RAGE signaling pathway in diabetic complications	91	1.628931462
20	HTLV-I infection	194	1.612144396
21	Bladder cancer	29	1.394263792
22	Vascular smooth muscle contraction	114	1.343415467
23	TNF signaling pathway	72	1.282307963
24	Amyotrophic lateral sclerosis (ALS)	36	1.229750618
25	D-Glutamine and D-glutamate metabolism	4	1.180851064
26	Cocaine addiction	42	1.179303568
27	Sulfur metabolism	9	1.171049088
28	MicroRNAs in cancer	262	1.157281824
29	Fc epsilon RI signaling pathway	61	1.142588348
30	Long-term depression	59	1.125250203
31	Chagas disease (American trypanosomiasis)	89	1.10760282
32	Salmonella infection	72	1.098184761
33	RNA degradation	18	1.083488434
34	Tight junction	125	1.077236311
35	Insulin signaling pathway	139	1.009797097
36	Leishmaniasis	50	0.985719996
37	Viral carcinogenesis	6	0.959459459
38	mTOR signaling pathway	144	0.94384078
39	Adherens junction	71	0.924425552
40	Oocyte meiosis	120	0.915631395
41	Bacterial invasion of epithelial cells	57	0.913453176
42	Tuberculosis	173	0.91194481
43	Vasopressin-regulated water reabsorption	22	0.897284788
44	RIG-I-like receptor signaling pathway	48	0.892891945

45	Inflammatory mediator regulation of TRP channels	91	0.878612581
46	Retrograde endocannabinoid signaling	59	0.852937341
47	Type II diabetes mellitus	47	0.843394635
48	Progesterone-mediated oocyte maturation	89	0.826138972
49	Thyroid hormone synthesis	46	0.822674589
50	Glutamatergic synapse	89	0.814505363

Table 3.9. Top 50 pathways ranked by T score comparing normal with hepatitis B-induced HCC samples based on the Pearson's correlation cutoff (-0.3).

		Gene Count(L)	T score
1	Circadian rhythm	31	4.602713001
2	FoxO signaling pathway	126	4.546560009
3	Hedgehog signaling pathway	47	4.280954331
4	Dorso-ventral axis formation	13	3.619026562
5	GnRH signaling pathway	85	3.515558749
6	Toll-like receptor signaling pathway	104	2.51299444
7	Wnt signaling pathway	137	2.398017415
8	MAPK signaling pathway	252	2.387515249
9	Rap1 signaling pathway	208	2.379274051
10	Epithelial cell signaling in Helicobacter pylori infection	37	2.082740987
11	Shigellosis	51	2.058152132
12	Renal cell carcinoma	57	2.043148412
13	Notch signaling pathway	48	1.947882308
14	Estrogen signaling pathway	89	1.895632946
15	Calcium signaling pathway	179	1.795913644
16	Hippo signaling pathway -multiple species	29	1.732928622
17	Gap junction	88	1.694405214
18	Thyroid cancer	28	1.671147819
19	AGE-RAGE signaling pathway in diabetic complications	91	1.628931462
20	HTLV-I infection	194	1.612144396
21	Bladder cancer	29	1.394263792
22	Vascular smooth muscle contraction	114	1.343415467
23	TNF signaling pathway	72	1.282307963
24	Amyotrophic lateral sclerosis (ALS)	36	1.229750618
25	D-Glutamine and D-glutamate metabolism	4	1.180851064
26	Cocaine addiction	42	1.179303568
27	Sulfur metabolism	9	1.171049088
28	MicroRNAs in cancer	262	1.157281824
29	Fc epsilon RI signaling pathway	61	1.142588348

30	Long-term depression	59	1.125250203
31	Chagas disease (American trypanosomiasis)	89	1.10760282
32	Salmonella infection	72	1.098184761
33	RNA degradation	18	1.083488434
34	Tight junction	125	1.077236311
35	Insulin signaling pathway	139	1.009797097
36	Leishmaniasis	50	0.985719996
37	Viral carcinogenesis	6	0.959459459
38	mTOR signaling pathway	144	0.94384078
39	Adherens junction	71	0.924425552
40	Oocyte meiosis	120	0.915631395
41	Bacterial invasion of epithelial cells	57	0.913453176
42	Tuberculosis	173	0.91194481
43	Vasopressin-regulated water reabsorption	22	0.897284788
44	RIG-I-like receptor signaling pathway	48	0.892891945
45	Inflammatory mediator regulation of TRP channels	91	0.878612581
46	Retrograde endocannabinoid signaling	59	0.852937341
47	Type II diabetes mellitus	47	0.843394635
48	Progesterone-mediated oocyte maturation	89	0.826138972
49	Thyroid hormone synthesis	46	0.822674589
50	Glutamatergic synapse	89	0.814505363

Table 3.10. Top 50 pathways ranked by T score comparing normal with hepatitis B-induced HCC samples based on the Pearson's correlation cutoff (-0.2).

		Gene Count(L)	T score
1	Circadian rhythm	31	4.602713001
2	FoxO signaling pathway	126	4.546560009
3	Hedgehog signaling pathway	47	4.280954331
4	Dorso-ventral axis formation	13	3.619026562
5	GnRH signaling pathway	85	3.515558749
6	Toll-like receptor signaling pathway	104	2.51299444
7	Wnt signaling pathway	137	2.398017415
8	MAPK signaling pathway	252	2.387515249
9	Rap1 signaling pathway	208	2.379274051
10	Epithelial cell signaling in Helicobacter pylori infection	37	2.082740987
11	Shigellosis	51	2.058152132
12	Renal cell carcinoma	57	2.043148412
13	Notch signaling pathway	48	1.947882308
14	Estrogen signaling pathway	89	1.895632946
15	Calcium signaling pathway	179	1.795913644

16	Hippo signaling pathway -multiple species	29	1.732928622
17	Gap junction	88	1.694405214
18	Thyroid cancer	28	1.671147819
19	AGE-RAGE signaling pathway in diabetic complications	91	1.628931462
20	HTLV-I infection	194	1.612144396
21	Bladder cancer	29	1.394263792
22	Vascular smooth muscle contraction	114	1.343415467
23	TNF signaling pathway	72	1.282307963
24	Amyotrophic lateral sclerosis (ALS)	36	1.229750618
25	D-Glutamine and D-glutamate metabolism	4	1.180851064
26	Cocaine addiction	42	1.179303568
27	Sulfur metabolism	9	1.171049088
28	MicroRNAs in cancer	262	1.157281824
29	Fc epsilon RI signaling pathway	61	1.142588348
30	Long-term depression	59	1.125250203
31	Chagas disease (American trypanosomiasis)	89	1.10760282
32	Salmonella infection	72	1.098184761
33	RNA degradation	18	1.083488434
34	Tight junction	125	1.077236311
35	Insulin signaling pathway	139	1.009797097
36	Leishmaniasis	50	0.985719996
37	Viral carcinogenesis	6	0.959459459
38	mTOR signaling pathway	144	0.94384078
39	Adherens junction	71	0.924425552
40	Oocyte meiosis	120	0.915631395
41	Bacterial invasion of epithelial cells	57	0.913453176
42	Tuberculosis	173	0.91194481
43	Vasopressin-regulated water reabsorption	22	0.897284788
44	RIG-I-like receptor signaling pathway	48	0.892891945
45	Inflammatory mediator regulation of TRP channels	91	0.878612581
46	Retrograde endocannabinoid signaling	59	0.852937341
47	Type II diabetes mellitus	47	0.843394635
48	Progesterone-mediated oocyte maturation	89	0.826138972
49	Thyroid hormone synthesis	46	0.822674589
50	Glutamatergic synapse	89	0.814505363

Table 3.11. Top 50 pathways ranked by T score comparing normal with hepatitis B-induced HCC samples based on the Pearson's correlation cutoff (-0.1).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	5.046943443

2	Rap1 signaling pathway	208	4.065776968
3	Hedgehog signaling pathway	47	3.836894713
4	Cocaine addiction	42	2.779786983
5	Circadian rhythm	31	2.594256419
6	Wnt signaling pathway	137	2.398017415
7	Hippo signaling pathway -multiple species	29	2.342737659
8	RNA degradation	18	1.950279181
9	Renal cell carcinoma	57	1.911332385
10	Notch signaling pathway	48	1.888855572
11	Vascular smooth muscle contraction	114	1.796469051
12	AGE-RAGE signaling pathway in diabetic complications	91	1.74328513
13	Shigellosis	51	1.741513343
14	MAPK signaling pathway	252	1.720852339
15	Bacterial invasion of epithelial cells	57	1.717291971
16	Steroid biosynthesis	20	1.622392948
17	mTOR signaling pathway	144	1.618051267
18	Estrogen signaling pathway	89	1.608441789
19	Thyroid cancer	28	1.606872903
20	Dorso-ventral axis formation	13	1.57348981
21	Calcium signaling pathway	179	1.532591743
22	MicroRNAs in cancer	262	1.506902683
23	Parkinson's disease	29	1.381536755
24	Progesterone-mediated oocyte maturation	89	1.343600254
25	Insulin resistance	94	1.285414079
26	Vasopressin-regulated water reabsorption	22	1.271153449
27	HTLV-I infection	194	1.217230571
28	Tight junction	125	1.212432605
29	Bladder cancer	29	1.203205273
30	Inflammatory bowel disease (IBD)	48	1.148826083
31	Leukocyte transendothelial migration	85	1.128676988
32	Epithelial cell signaling in Helicobacter pylori infection	37	1.109696624
33	Insulin signaling pathway	139	1.055000849
34	Adipocytokine signaling pathway	63	1.046300074
35	Amyotrophic lateral sclerosis (ALS)	36	0.983800494
36	GnRH signaling pathway	85	0.966264685
37	Breast cancer	143	0.943208322
38	Oocyte meiosis	120	0.927522712
39	Adherens junction	71	0.904756923
40	Long-term potentiation	67	0.873254968
41	Axon guidance	167	0.867502498
42	Cytosolic DNA-sensing pathway	21	0.862616564

43	ECM-receptor interaction	81	0.858243412
44	Toll-like receptor signaling pathway	104	0.851435884
45	Long-term depression	59	0.834863053
46	alpha-Linolenic acid metabolism	25	0.815135135
47	Pancreatic cancer	65	0.80634025
48	Gap junction	88	0.799606955
49	Signaling pathways regulating pluripotency of stem cells	112	0.783109171
50	Alzheimer's disease	48	0.771233168

Table 3.12. Top 50 pathways ranked by T score comparing normal with hepatitis C-induced HCC samples based on the Pearson's correlation cutoff (-0.3).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	5.046943443
2	Rap1 signaling pathway	208	4.065776968
3	Hedgehog signaling pathway	47	3.836894713
4	Cocaine addiction	42	2.779786983
5	Circadian rhythm	31	2.594256419
6	Wnt signaling pathway	137	2.398017415
7	Hippo signaling pathway -multiple species	29	2.342737659
8	RNA degradation	18	1.950279181
9	Renal cell carcinoma	57	1.911332385
10	Notch signaling pathway	48	1.888855572
11	Vascular smooth muscle contraction	114	1.796469051
12	AGE-RAGE signaling pathway in diabetic complications	91	1.74328513
13	Shigellosis	51	1.741513343
14	MAPK signaling pathway	252	1.720852339
15	Bacterial invasion of epithelial cells	57	1.717291971
16	Steroid biosynthesis	20	1.622392948
17	mTOR signaling pathway	144	1.618051267
18	Estrogen signaling pathway	89	1.608441789
19	Thyroid cancer	28	1.606872903
20	Dorso-ventral axis formation	13	1.57348981
21	Calcium signaling pathway	179	1.532591743
22	MicroRNAs in cancer	262	1.506902683
23	Parkinson's disease	29	1.381536755
24	Progesterone-mediated oocyte maturation	89	1.343600254
25	Insulin resistance	94	1.285414079
26	Vasopressin-regulated water reabsorption	22	1.271153449
27	HTLV-I infection	194	1.217230571
28	Tight junction	125	1.212432605

29	Bladder cancer	29	1.203205273
30	Inflammatory bowel disease (IBD)	48	1.148826083
31	Leukocyte transendothelial migration	85	1.128676988
32	Epithelial cell signaling in Helicobacter pylori infection	37	1.109696624
33	Insulin signaling pathway	139	1.055000849
34	Adipocytokine signaling pathway	63	1.046300074
35	Amyotrophic lateral sclerosis (ALS)	36	0.983800494
36	GnRH signaling pathway	85	0.966264685
37	Breast cancer	143	0.943208322
38	Oocyte meiosis	120	0.927522712
39	Adherens junction	71	0.904756923
40	Long-term potentiation	67	0.873254968
41	Axon guidance	167	0.867502498
42	Cytosolic DNA-sensing pathway	21	0.862616564
43	ECM-receptor interaction	81	0.858243412
44	Toll-like receptor signaling pathway	104	0.851435884
45	Long-term depression	59	0.834863053
46	alpha-Linolenic acid metabolism	25	0.815135135
47	Pancreatic cancer	65	0.80634025
48	Gap junction	88	0.799606955
49	Signaling pathways regulating pluripotency of stem cells	112	0.783109171
50	Alzheimer's disease	48	0.771233168

Table 3.13. Top 50 pathways ranked by T score comparing normal with hepatitis C-induced HCC samples based on the Pearson's correlation cutoff (-0.2).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	5.046943443
2	Rap1 signaling pathway	208	4.065776968
3	Hedgehog signaling pathway	47	3.836894713
4	Cocaine addiction	42	2.779786983
5	Circadian rhythm	31	2.594256419
6	Wnt signaling pathway	137	2.398017415
7	Hippo signaling pathway -multiple species	29	2.342737659
8	RNA degradation	18	1.950279181
9	Renal cell carcinoma	57	1.911332385
10	Notch signaling pathway	48	1.888855572
11	Vascular smooth muscle contraction	114	1.796469051
12	AGE-RAGE signaling pathway in diabetic complications	91	1.74328513
13	Shigellosis	51	1.741513343

14	MAPK signaling pathway	252	1.720852339
15	Bacterial invasion of epithelial cells	57	1.717291971
16	Steroid biosynthesis	20	1.622392948
17	mTOR signaling pathway	144	1.618051267
18	Estrogen signaling pathway	89	1.608441789
19	Thyroid cancer	28	1.606872903
20	Dorso-ventral axis formation	13	1.57348981
21	Calcium signaling pathway	179	1.532591743
22	MicroRNAs in cancer	262	1.506902683
23	Parkinson's disease	29	1.381536755
24	Progesterone-mediated oocyte maturation	89	1.343600254
25	Insulin resistance	94	1.285414079
26	Vasopressin-regulated water reabsorption	22	1.271153449
27	HTLV-I infection	194	1.217230571
28	Tight junction	125	1.212432605
29	Bladder cancer	29	1.203205273
30	Inflammatory bowel disease (IBD)	48	1.148826083
31	Leukocyte transendothelial migration	85	1.128676988
32	Epithelial cell signaling in Helicobacter pylori infection	37	1.109696624
33	Insulin signaling pathway	139	1.055000849
34	Adipocytokine signaling pathway	63	1.046300074
35	Amyotrophic lateral sclerosis (ALS)	36	0.983800494
36	GnRH signaling pathway	85	0.966264685
37	Breast cancer	143	0.943208322
38	Oocyte meiosis	120	0.927522712
39	Adherens junction	71	0.904756923
40	Long-term potentiation	67	0.873254968
41	Axon guidance	167	0.867502498
42	Cytosolic DNA-sensing pathway	21	0.862616564
43	ECM-receptor interaction	81	0.858243412
44	Toll-like receptor signaling pathway	104	0.851435884
45	Long-term depression	59	0.834863053
46	alpha-Linolenic acid metabolism	25	0.815135135
47	Pancreatic cancer	65	0.80634025
48	Gap junction	88	0.799606955
49	Signaling pathways regulating pluripotency of stem cells	112	0.783109171
50	Alzheimer's disease	48	0.771233168

Table 3.14. Top 50 pathways ranked by T score comparing normal with hepatitis C-induced HCC samples based on the Pearson's correlation cutoff (-0.1).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	3.81448289
2	Cocaine addiction	42	3.622146675
3	Hedgehog signaling pathway	47	3.071150497
4	Rap1 signaling pathway	208	2.442721359
5	Circadian rhythm	31	2.426885037
6	Fatty acid biosynthesis	13	2.385487428
7	Estrogen signaling pathway	89	2.363686878
8	RNA degradation	18	2.22115129
9	Calcium signaling pathway	179	2.059235544
10	Hippo signaling pathway -multiple species	29	2.051175853
11	Notch signaling pathway	48	1.829828835
12	Glutamatergic synapse	89	1.820659046
13	Morphine addiction	54	1.665147311
14	alpha-Linolenic acid metabolism	25	1.63027027
15	Vascular smooth muscle contraction	114	1.617006058
16	Retrograde endocannabinoid signaling	59	1.578841462
17	Shigellosis	51	1.551530069
18	Wnt signaling pathway	137	1.539779604
19	MAPK signaling pathway	252	1.457294704
20	Pancreatic secretion	30	1.382495611
21	Long-term depression	59	1.379338958
22	Vasopressin-regulated water reabsorption	22	1.345927182
23	HTLV-I infection	194	1.327880745
24	Insulin resistance	94	1.325244579
25	GABAergic synapse	66	1.285849245
26	mTOR signaling pathway	144	1.245297312
27	Thyroid hormone synthesis	46	1.22254989
28	Huntington's disease	27	1.195229442
29	Bacterial invasion of epithelial cells	57	1.132681938
30	Maturity onset diabetes of the young	24	1.101228136
31	Circadian entrainment	96	1.053790085
32	MicroRNAs in cancer	262	1.053638972
33	Gap junction	88	1.028066085
34	Inflammatory bowel disease (IBD)	48	1.00344928
35	Progesterone-mediated oocyte maturation	89	0.983890456
36	AGE-RAGE signaling pathway in diabetic complications	91	0.972996351
37	GnRH signaling pathway	85	0.925147039
38	Long-term potentiation	67	0.894620948
39	Dopaminergic synapse	124	0.893777542

40	Epithelial cell signaling in Helicobacter pylori infection	37	0.880104219
41	Oocyte meiosis	120	0.879957445
42	Chagas disease (American trypanosomiasis)	89	0.836077118
43	Aldosterone synthesis and secretion	65	0.82394993
44	Toll-like receptor signaling pathway	104	0.821858671
45	Melanogenesis	101	0.820322345
46	Alzheimer's disease	48	0.807137423
47	Insulin secretion	54	0.792002705
48	Renal cell carcinoma	57	0.790896159
49	Transcriptional misregulation in cancer	19	0.789473684
50	Valine, leucine and isoleucine degradation	48	0.783476417

Table 3.15. Top 50 pathways ranked by T score comparing normal with alcohol-induced HCC samples based on the Pearson's correlation cutoff (-0.3).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	6.253538533
2	Cocaine addiction	42	5.222630089
3	Hedgehog signaling pathway	47	4.694720306
4	Rap1 signaling pathway	208	3.85374078
5	Circadian rhythm	31	3.682170401
6	Estrogen signaling pathway	89	3.520828277
7	Calcium signaling pathway	179	3.367955103
8	Hippo signaling pathway -multiple species	29	3.270793928
9	RNA degradation	18	2.762895507
10	Glutamatergic synapse	89	2.68307649
11	Wnt signaling pathway	137	2.650440301
12	alpha-Linolenic acid metabolism	25	2.445405405
13	Notch signaling pathway	48	2.420096201
14	Morphine addiction	54	2.417885137
15	Shigellosis	51	2.406454801
16	Retrograde endocannabinoid signaling	59	2.395483597
17	Fatty acid biosynthesis	13	2.385487428
18	Vascular smooth muscle contraction	114	2.382841389
19	Dorso-ventral axis formation	13	2.202885734
20	Pancreatic secretion	30	2.172493102
21	Long-term depression	59	2.141605224
22	MAPK signaling pathway	252	2.141135101
23	Renal cell carcinoma	57	2.109056425
24	Vasopressin-regulated water reabsorption	22	2.093664505
25	AGE-RAGE signaling pathway in diabetic complications	91	2.027075732

26	Insulin resistance	94	1.980597638
27	HTLV-I infection	194	1.959119287
28	mTOR signaling pathway	144	1.939310767
29	Huntington's disease	27	1.930755252
30	Thyroid hormone synthesis	46	1.828730001
31	GABAergic synapse	66	1.807643141
32	GnRH signaling pathway	85	1.685823494
33	Bacterial invasion of epithelial cells	57	1.680753844
34	Maturity onset diabetes of the young	24	1.651842204
35	Circadian entrainment	96	1.580685128
36	Gap junction	88	1.580175649
37	Progesterone-mediated oocyte maturation	89	1.551298787
38	Epithelial cell signaling in Helicobacter pylori infection	37	1.492350632
39	MicroRNAs in cancer	262	1.491486361
40	Salmonella infection	72	1.453479831
41	Long-term potentiation	67	1.426557728
42	Toll-like receptor signaling pathway	104	1.41340293
43	Insulin signaling pathway	139	1.360570584
44	Tight junction	125	1.350305952
45	Dopaminergic synapse	124	1.329235264
46	Oocyte meiosis	120	1.308044851
47	Inflammatory bowel disease (IBD)	48	1.294202886
48	Melanogenesis	101	1.282685848
49	Inflammatory mediator regulation of TRP channels	91	1.258568268
50	Insulin secretion	54	1.257426218

Table 3.16. Top 50 pathways ranked by T score comparing normal with alcohol-induced HCC samples based on the Pearson's correlation cutoff (-0.2).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	6.253538533
2	Cocaine addiction	42	5.222630089
3	Hedgehog signaling pathway	47	4.694720306
4	Rap1 signaling pathway	208	3.85374078
5	Circadian rhythm	31	3.682170401
6	Estrogen signaling pathway	89	3.520828277
7	Calcium signaling pathway	179	3.367955103
8	Hippo signaling pathway -multiple species	29	3.270793928
9	RNA degradation	18	2.762895507

10	Glutamatergic synapse	89	2.68307649
11	Wnt signaling pathway	137	2.650440301
12	alpha-Linolenic acid metabolism	25	2.445405405
13	Notch signaling pathway	48	2.420096201
14	Morphine addiction	54	2.417885137
15	Shigellosis	51	2.406454801
16	Retrograde endocannabinoid signaling	59	2.395483597
17	Fatty acid biosynthesis	13	2.385487428
18	Vascular smooth muscle contraction	114	2.382841389
19	Dorso-ventral axis formation	13	2.202885734
20	Pancreatic secretion	30	2.172493102
21	Long-term depression	59	2.141605224
22	MAPK signaling pathway	252	2.141135101
23	Renal cell carcinoma	57	2.109056425
24	Vasopressin-regulated water reabsorption	22	2.093664505
25	AGE-RAGE signaling pathway in diabetic complications	91	2.027075732
26	Insulin resistance	94	1.980597638
27	HTLV-I infection	194	1.959119287
28	mTOR signaling pathway	144	1.939310767
29	Huntington's disease	27	1.930755252
30	Thyroid hormone synthesis	46	1.828730001
31	GABAergic synapse	66	1.807643141
32	GnRH signaling pathway	85	1.685823494
33	Bacterial invasion of epithelial cells	57	1.680753844
34	Maturity onset diabetes of the young	24	1.651842204
35	Circadian entrainment	96	1.580685128
36	Gap junction	88	1.580175649
37	Progesterone-mediated oocyte maturation	89	1.551298787
38	Epithelial cell signaling in Helicobacter pylori infection	37	1.492350632
39	MicroRNAs in cancer	262	1.491486361
40	Salmonella infection	72	1.453479831
41	Long-term potentiation	67	1.426557728
42	Toll-like receptor signaling pathway	104	1.41340293
43	Insulin signaling pathway	139	1.360570584
44	Tight junction	125	1.350305952
45	Dopaminergic synapse	124	1.329235264
46	Oocyte meiosis	120	1.308044851
47	Inflammatory bowel disease (IBD)	48	1.294202886
48	Melanogenesis	101	1.282685848
49	Inflammatory mediator regulation of TRP channels	91	1.258568268

50	Insulin secretion	54	1.257426218

Table 3.17. Top 50 pathways ranked by T score comparing normal with alcohol-induced HCC samples based on the Pearson's correlation cutoff (-0.1).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	0.285933
2	RNA degradation	18	0.270872
3	Bladder cancer	29	0.133689
4	MicroRNAs in cancer	262	0.111402
5	Hedgehog signaling pathway	47	0.107228
6	Circadian rhythm	31	0.083686
7	Amyotrophic lateral sclerosis (ALS)	36	0.081983
8	Platinum drug resistance	41	0.080392
9	p53 signaling pathway	68	0.066399
10	Shigellosis	51	0.063328
11	Phototransduction	27	0.062432
12	Cell cycle	124	0.057685
13	Hippo signaling pathway -multiple species	29	0.055437
14	Transcriptional misregulation in cancer	19	0.052632
15	Pancreatic cancer	65	0.047432
16	HTLV-I infection	194	0.041036
17	AGE-RAGE signaling pathway in diabetic complications	91	0.040542
18	Bacterial invasion of epithelial cells	57	0.036538
19	Small cell lung cancer	83	0.03451
20	Pathogenic Escherichia coli infection	40	0.03419
21	Hepatitis B	134	0.033589
22	Phagosome	32	0.033142
23	HIF-1 signaling pathway	102	0.032685
24	Galactose metabolism	28	0.032464
25	Salmonella infection	72	0.0323
26	Chronic myeloid leukemia	73	0.032277
27	Breast cancer	143	0.032064
28	EGFR tyrosine kinase inhibitor resistance	81	0.029106
29	Non-small cell lung cancer	54	0.02828
30	Glioma	66	0.028163
31	Melanoma	69	0.027454
32	Starch and sucrose metabolism	36	0.026393
33	PI3K-Akt signaling pathway	340	0.026082
34	Adipocytokine signaling pathway	63	0.02552
35	Wnt signaling pathway	137	0.025242

36	Sphingolipid metabolism	47	0.02522
37	Oocyte meiosis	120	0.023783
38	Insulin signaling pathway	139	0.022193
39	Progesterone-mediated oocyte maturation	89	0.021561
40	Apoptosis	133	0.020906
41	Tuberculosis	173	0.019046
42	Gap junction	88	0.019038
43	Insulin resistance	94	0.01884
44	Pathways in cancer	310	0.018686
45	Protein processing in endoplasmic reticulum	51	0.018554
46	MAPK signaling pathway	252	0.017528
47	Glycerophospholipid metabolism	94	0.017197
48	NF-kappa B signaling pathway	81	0.017163
49	Central carbon metabolism in cancer	63	0.017052
50	Leukocyte transendothelial migration	85	0.016573

Table 3.18. Top 50 pathways ranked by T score comparing normal with HCC samples based on the Pearson's correlation cutoff (-0.4).

		Gene Count(L)	T score
1	FoxO signaling pathway	126	1.366805
2	Hedgehog signaling pathway	47	1.141043
3	Circadian rhythm	31	1.087914
4	Dorso-ventral axis formation	13	0.786745
5	Rap1 signaling pathway	208	0.666197
6	Wnt signaling pathway	137	0.6563
7	RNA degradation	18	0.650093
8	Oxidative phosphorylation	47	0.619513
9	Renal cell carcinoma	57	0.593172
10	Bladder cancer	29	0.534758
11	Ribosome biogenesis in eukaryotes	3	0.486486
12	Shigellosis	51	0.474958
13	Toll-like receptor signaling pathway	104	0.473235
14	MAPK signaling pathway	252	0.472747
15	HTLV-I infection	194	0.463854
16	GnRH signaling pathway	85	0.411176
17	Amyotrophic lateral sclerosis (ALS)	36	0.409917
18	MicroRNAs in cancer	262	0.389905
19	Hippo signaling pathway -multiple species	29	0.38806
20	Vasopressin-regulated water reabsorption	22	0.373869
21	Estrogen signaling pathway	89	0.369246

22	Parkinson's disease	29	0.345384
23	Fatty acid biosynthesis	13	0.340784
24	Sphingolipid metabolism	47	0.327862
25	AGE-RAGE signaling pathway in diabetic complications	91	0.324332
26	Salmonella infection	72	0.322996
27	Viral carcinogenesis	6	0.31982
28	Legionellosis	40	0.316216
29	Epithelial cell signaling in Helicobacter pylori infection	37	0.306123
30	Platinum drug resistance	41	0.294772
31	Sulfur metabolism	9	0.292762
32	Bacterial invasion of epithelial cells	57	0.292305
33	Antigen processing and presentation	62	0.273428
34	Thyroid cancer	28	0.2571
35	Cocaine addiction	42	0.252708
36	Notch signaling pathway	48	0.236107
37	RIG-I-like receptor signaling pathway	48	0.22424
38	Insulin signaling pathway	139	0.221931
39	Regulation of autophagy	19	0.217501
40	Nitrogen metabolism	4	0.2173
41	TNF signaling pathway	72	0.204284
42	Inflammatory mediator regulation of TRP channels	91	0.202757
43	Adherens junction	71	0.196686
44	Fc epsilon RI signaling pathway	61	0.195855
45	Thyroid hormone synthesis	46	0.19357
46	Pyrimidine metabolism	105	0.188895
47	p53 signaling pathway	68	0.185916
48	Vascular smooth muscle contraction	114	0.182394
49	Tuberculosis	173	0.180941
50	HIF-1 signaling pathway	102	0.179767

Table 3.19. Top 50 pathways ranked by T score comparing normal with hepatitis B-induced HCC samples based on the Pearson's correlation cutoff (-0.4).

		Gene Count(L)	T score
1	Hedgehog signaling pathway	47	0.589755
2	RNA degradation	18	0.433395
3	Bladder cancer	29	0.401068
4	FoxO signaling pathway	126	0.366039
5	Fatty acid biosynthesis	13	0.340784
6	Circadian rhythm	31	0.334743

7	Lipoic acid metabolism	3	0.303191
8	Steroid biosynthesis	20	0.284026
9	Porphyrin and chlorophyll metabolism	39	0.26793
10	Homologous recombination	18	0.241308
11	Nitrogen metabolism	4	0.2173
12	Shigellosis	51	0.189983
13	Parkinson's disease	29	0.172692
14	Amyotrophic lateral sclerosis (ALS)	36	0.163967
15	Rap1 signaling pathway	208	0.158618
16	MicroRNAs in cancer	262	0.153177
17	Wnt signaling pathway	137	0.151454
18	Sulfur metabolism	9	0.146381
19	Platinum drug resistance	41	0.133987
20	Drug metabolism - cytochrome P450	70	0.13375
21	Fanconi anemia pathway	40	0.133169
22	Butanoate metabolism	27	0.114959
23	Viral myocarditis	26	0.113768
24	Hippo signaling pathway -multiple species	29	0.110874
25	Valine, leucine and isoleucine degradation	48	0.109783
26	Bacterial invasion of epithelial cells	57	0.109614
27	Legionellosis	40	0.105405
28	Pathogenic Escherichia coli infection	40	0.10257
29	Sphingolipid metabolism	47	0.100881
30	Cell cycle	124	0.096141
31	Fatty acid degradation	42	0.092897
32	Vascular smooth muscle contraction	114	0.091197
33	MAPK signaling pathway	252	0.087642
34	Ascorbate and aldarate metabolism	21	0.084129
35	Lysine degradation	55	0.082721
36	AGE-RAGE signaling pathway in diabetic complications	91	0.081083
37	TGF-beta signaling pathway	73	0.080692
38	p53 signaling pathway	68	0.079678
39	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	10	0.077027
40	Tuberculosis	173	0.076186
41	Folate biosynthesis	14	0.071429
42	Pancreatic cancer	65	0.071148
43	HTLV-I infection	194	0.068393
44	Phagosome	32	0.066284

45	Renal cell carcinoma	57	0.065908
46	Natural killer cell mediated cytotoxicity	134	0.064933
47	Salmonella infection	72	0.064599
48	Colorectal cancer	49	0.064341
49	Thyroid cancer	28	0.064275
50	Breast cancer	143	0.064129

Table 3.20. Top 50 pathways ranked by T score comparing normal with hepatitis C-induced HCC samples based on the Pearson's correlation cutoff (-0.4).

		Gene Count(L)	T score
1	Fatty acid biosynthesis	13	0.340784
2	RNA degradation	18	0.270872
3	FoxO signaling pathway	126	0.21445
4	Bladder cancer	29	0.133689
5	Phototransduction	27	0.124864
6	Hedgehog signaling pathway	47	0.107228
7	Circadian rhythm	31	0.083686
8	Amyotrophic lateral sclerosis (ALS)	36	0.081983
9	Platinum drug resistance	41	0.080392
10	Shigellosis	51	0.063328
11	MicroRNAs in cancer	262	0.055701
12	Hippo signaling pathway -multiple species	29	0.055437
13	Propanoate metabolism	32	0.053321
14	AGE-RAGE signaling pathway in diabetic complications	91	0.040542
15	Bacterial invasion of epithelial cells	57	0.036538
16	Pathogenic Escherichia coli infection	40	0.03419
17	Phagosome	32	0.033142
18	HIF-1 signaling pathway	102	0.032685
19	Salmonella infection	72	0.0323
20	EGFR tyrosine kinase inhibitor resistance	81	0.029106
21	HTLV-I infection	194	0.027357
22	Wnt signaling pathway	137	0.025242
23	Sphingolipid metabolism	47	0.02522
24	Pyruvate metabolism	39	0.023651
25	Insulin signaling pathway	139	0.022193
26	Apoptosis	133	0.020906
27	Cell cycle	124	0.019228
28	Tuberculosis	173	0.019046
29	Protein processing in endoplasmic reticulum	51	0.018554

30	MAPK signaling pathway	252	0.017528
31	NF-kappa B signaling pathway	81	0.017163
32	Central carbon metabolism in cancer	63	0.017052
33	Natural killer cell mediated cytotoxicity	134	0.016233
34	Colorectal cancer	49	0.016085
35	RNA transport	133	0.015116
36	AMPK signaling pathway	97	0.014694
37	PI3K-Akt signaling pathway	340	0.01449
38	Hepatitis C	97	0.014287
39	Endocrine resistance	95	0.013688
40	Hepatitis B	134	0.013436
41	p53 signaling pathway	68	0.01328
42	Glucagon signaling pathway	86	0.013129
43	Chronic myeloid leukemia	73	0.012911
44	TNF signaling pathway	72	0.012768
45	Circadian entrainment	96	0.012545
46	Oocyte meiosis	120	0.011891
47	Pancreatic cancer	65	0.011858
48	Prostate cancer	87	0.010359
49	Pathways in cancer	310	0.009343
50	Hippo signaling pathway	151	0.009231

Table 3.21. Top 50 pathways ranked by T score comparing normal with alcohol-induced HCC samples based on the Pearson's correlation cutoff (-0.4).

Chapter 4: Conclusion, Discussion and Next Steps

Complex diseases such as cancer are usually result from a combination of environmental factors and one or several biological pathways consisting of sets of genes. Each biological pathways exerts its own function by delivering signaling through the gene network. The interactions and reactions of genes and their products constitute a biological pathway. The activation and inhibition of biological pathways directly influence on the onset and development of complex diseases. Hence, it is fairly important to detect disease-related biological pathways.

A biological pathway might be influenced by some aberrant genes in this pathway or some aberrant regulators, such as miRNAs, that regulate this pathway. In this dissertation, we investigate how these two layers of components, genes and miRNAs, cause the aberration of biological pathways.

It is well known that aberrantly expressed genes in the pathway and/or aberrantly expressed miRNAs regulating the pathways can trigger dysregulation of the biological pathways. There are many studies on this area. The interplay of genes, miRNAs and biological pathways has been increasing active over the past decade. Some methods have been created to identify functional networks across different phenotypes from changes in gene expression, as well as some methods have been created to identify miRNA regulatory networks across different phenotypes from changes in miRNA expression. On the other hand, altered gene-gene connections in biological pathways and/or altered miRNA-mRNA connections might also impact the activity of biological pathways. In this dissertation, we discuss how these altered gene-gene connections (Chapter 1) and altered miRNA-mRNA connections (Chapter 2) aberrantly influence the activity of biological pathways and their association with disease.

In Chapter 1, we propose a method called PoTRA, Pathways of Topological Rank Analysis, to detect pathways involved in cancer. PoTRA is motivated by that the loss of connectivity is a common topological trait of cancer networks, and the prior knowledge that a normal biological network is a scale-free network whose degree distribution follows a power law where a small number of nodes are hubs and a large number of nodes are non-hubs. From normal to cancer state, the process of the network losing connectivity could be the process of disrupting the scale-free structure of the network, that is to say, the number of hub genes might be altered in cancer compared to that in normal samples. Hence, it is hypothesized that if the number of hub genes is different in a pathway between normal and cancer, this pathway might be involved in cancer. Based on this hypothesis, we propose to detect pathways involved in cancer by testing the pathways with altered number of hub genes between normal and cancer samples.

Thus, the PoTRA method focuses on topological ranks of genes in each pathway, so we use the Google search PageRank algorithm to compute the relative topological ranks of genes in each pathway across different phenotypes, and then select the hub genes for each pathway, then detects pathways with significantly altered number of hub genes between normal and cancer samples. For the testing step, we use Fisher's exact test to estimate if the number of hub genes in each pathway is altered between normal and cancer. We apply PoTRA to HCC and three subtypes of HCC (hepatitis B, hepatitis C, alcohol). We discover many HCC-relevant pathways generally and HCC subtype-relevant pathways specifically.

In Chapter 2, we propose a new PageRank-based method, called miR2Pathway, to quantifying the differential effects of miRNAs on the activity of a biological pathway and rank disease risk of rewired miRNA-mediated biological pathways. It is motivated that a miRNA might differentially regulate genes between normal and cancer, and these

differential regulations could aberrantly influence activity of biological pathways. Namely, there are two cases. In case 1, a miRNA might regulate several genes in normal tissue, while this miRNA might lose these regulatory connections in cancer tissue; in case 2, this miRNA regulates the same number of non-hub genes in this pathway in normal tissue, while it loses the regulatory connections in tumor tissue. Our hypothesis is that this miRNA has a larger differential influence on the activity of the pathway in case 1 than case 2. We use PageRank to compute the topological influence (PageRank scores) of the targeted genes in this differential network, which quantifies the topological influence of the genes that are differentially targeted by a miRNA on the activity of this pathway. Then we can compute the sum of PageRank scores of genes targeted by the miRNA in the differential network, which estimates the total differential influence of a miRNA on the activity of this pathway. Then, the same procedure is repeated for all miRNAs, so that we can obtain a corresponding sum of PageRank scores for each miRNA. For a specific pathway, we estimate the total differential influence of all the miRNAs on this pathway through summing up all the sums corresponding to all the miRNAs. Namely, the total differential influence of all the miRNAs on this pathway reflects the degree of miRNA-mediated dysregulation of this pathway. We do this for each pathway. Finally, we rank all the pathways by the degree of miRNA-mediated dysregulation scores. We apply miR2Pathway to HCC and the three subtypes of HCC. Very interestingly, we also discover many HCC-relevant pathways generally and HCC subtype-relevant pathways specifically.

For PoTRA, the hypothesis of our study is based on the fact that the loss of connectivity is a common topological trait of cancer networks. However, we still cannot well understand if this trait is characteristic of other complex diseases. Hence, we have to be careful about the applicability of PoTRA to other diseases. Although PoTRA is motivated

by work on cancer, it could apply to other complex diseases as well. This direction needs to be further explored. For miR2Pathway, miR2Pathway is applied to pre-defined biological pathways, from the well-curated KEGG pathway database. To extend this application, we also can apply miR2Pathway to any set of genes of interest, such as functional gene networks. Additionally, miR2Pathway might also be used to estimate differential influence of other regulatory factors, e.g., transcriptional factors (TFs), on biological pathways. For another point, miR2Pathway is used to rank the degree of miRNA-mediated dysregulation in our study. It is seen that there is no cutoff in it. Therefore, how to set up the cutoff level will become an important and interesting direction to explore in the future.

Reference

Albert R. 2005. Scale-free networks in cell biology. *Journal of Cell Science* 118:4947–4957. DOI: 10.1242/jcs.02714.

Albert R., Barabási A-L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74:47–97. DOI: 10.1103/RevModPhys.74.47.

Anglani R., Creanza TM., Liuzzi VC., Piepoli A., Panza A., Andriulli A., Ancona N. 2014. Loss of Connectivity in Cancer Co-Expression Networks. *PLOS ONE* 9:e87075. DOI: 10.1371/journal.pone.0087075.

Appert-Collin A., Hubert P., Crémel G., Bennasroune A. 2015. Role of ErbB Receptors in Cancer Cell Migration and Invasion. *Frontiers in Pharmacology* 6. DOI: 10.3389/fphar.2015.00283.

Aqil M., Mallik S., Bandyopadhyay S., Maulik U., Jameel S. 2015.Transcriptomic Analysis of mRNAs in Human Monocytic Cells Expressing the HIV-1 Nef Protein and Their Exosomes. *Available at https://www.hindawi.com/journals/bmri/2015/492395/* (accessed September 23, 2017). DOI: 10.1155/2015/492395.

Aqil M., Naqvi AR., Mallik S., Bandyopadhyay S., Maulik U., Jameel S. 2014. The HIV Nef protein modulates cellular and exosomal miRNA profiles in human monocytic cells. *Journal of Extracellular Vesicles* 3. DOI: 10.3402/jev.v3.23129.

Backes C., Kehl T., Stöckel D., Fehlmann T., Schneider L., Meese E., Lenhof H-P., Keller A. 2017. miRPathDB: a new dictionary on microRNAs and target pathways. *Nucleic Acids Research* 45:D90–D96. DOI: 10.1093/nar/gkw926.

Backes C., Khaleeq QT., Meese E., Keller A. 2016. miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Research* 44:W110-116. DOI: 10.1093/nar/gkw345.

Bailey CL. 2009. The Roles of Rap1 in Cancer Metastasis and Pancreatic Islet Beta Cell Function.

Bandiera S., Pfeffer S., Baumert TF., Zeisel MB. 2015. miR-122 – A key factor and therapeutic target in liver disease. *Journal of Hepatology* 62:448–457. DOI: 10.1016/j.jhep.2014.10.004.

Barthel SR., Medvedev R., Heinrich T., Büchner SM., Kettern N., Hildt E. 2016. Hepatitis B virus inhibits insulin receptor signaling and impairs liver regeneration via intracellular retention of the insulin receptor. *Cellular and molecular life sciences: CMLS* 73:4121–4140. DOI: 10.1007/s00018-016-2259-1.

Beasley RP. 1988. Hepatitis B virus. The major etiology of hepatocellular carcinoma. *Cancer* 61:1942–1956.

Ben-Hamo R., Efroni S. 2015. MicroRNA regulation of molecular pathways as a generic mechanism and as a core disease phenotype. *Oncotarget* 6:1594–1604. DOI: 10.18632/oncotarget.2734.

Benjamini Y., Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57:289–300.

Bishayee A. 2014. The role of inflammation and liver cancer. *Advances in Experimental Medicine and Biology* 816:401–435. DOI: 10.1007/978-3-0348-0837-8_16.

Blakeman V., Williams JL., Meng Q-J., Streuli CH. 2016. Circadian clocks and breast cancer. *Breast cancer research: BCR* 18:89. DOI: 10.1186/s13058-016-0743-z.

Bos JL., de Rooij J., Reedquist KA. 2001. Rap1 signalling: adhering to new models. *Nature Reviews Molecular Cell Biology* 2:369–377. DOI: 10.1038/35073073.

Bose SK. 2014. Hepatitis C virus infection and insulin resistance. *World J Diabetes* 5:52. DOI: 10.4239/wjd.v5.i1.52.

Bourdakou MM., Athanasiadis EI., Spyrou GM. 2016. Discovering gene re-ranking efficiency and conserved gene-gene relationships derived from gene co-expression network analysis on breast cancer data. *Scientific Reports* 6:20518. DOI: 10.1038/srep20518.

Burotto M., Chiou VL., Lee J-M., Kohn EC. 2014. The MAPK pathway across different malignancies: A new perspective. *Cancer* 120:3446–3456. DOI: 10.1002/cncr.28864.

Calin GA., Croce CM. 2006. MicroRNA signatures in human cancers. *Nature Reviews. Cancer* 6:857–866. DOI: 10.1038/nrc1997.

Capaccione KM., Pine SR. 2013. The Notch signaling pathway as a mediator of tumor survival. *Carcinogenesis* 34:1420–1430. DOI: 10.1093/carcin/bgt127.

Chang J., Nicolas E., Marks D., Sander C., Lerro A., Buendia MA., Xu C., Mason WS., Moloshok T., Bort R., Zaret KS., Taylor JM. 2004. miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA biology* 1:106–113.

Chédotal A., Kerjan G., Moreau-Fauvarque C. 2005. The brain within the tumor: new roles for axon guidance molecules in cancers. *Cell Death and Differentiation* 12:1044–1056. DOI: 10.1038/sj.cdd.4401707.

Chen C-L., Yang H-I., Yang W-S., Liu C-J., Chen P-J., You S-L., Wang L-Y., Sun C-A., Lu S-N., Chen D-S., Chen C-J. 2008. Metabolic factors and risk of hepatocellular carcinoma by chronic hepatitis B/C infection: a follow-up study in Taiwan. *Gastroenterology* 135:111–121. DOI: 10.1053/j.gastro.2008.03.073.

Cho H-S., Leahy DJ. 2002. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science (New York, N.Y.)* 297:1330–1333. DOI: 10.1126/science.1074611.

Choi Y., Kendziorski C. 2009. Statistical methods for gene set co-expression analysis. *Bioinformatics (Oxford, England)* 25:2780–2786. DOI: 10.1093/bioinformatics/btp502.

Chopin V., Lagadec C., Toillon R-A., Le Bourhis X. 2016. Neurotrophin signaling in cancer stem cells. *Cellular and molecular life sciences: CMLS* 73:1859–1870. DOI: 10.1007/s00018-016-2156-7.

Cimmino A., Calin GA., Fabbri M., Iorio MV., Ferracin M., Shimizu M., Wojcik SE., Aqeilan RI., Zupo S., Dono M., Rassenti L., Alder H., Volinia S., Liu C-G., Kipps TJ., Negrini M., Croce CM. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proceedings of the National Academy of Sciences of the United States of America* 102:13944–13949. DOI: 10.1073/pnas.0506654102.

Cirera S., Birck M., Busk PK., Fredholm M. 2010. Expression profiles of miRNA-122 and its target CAT1 in minipigs (Sus scrofa) fed a high-cholesterol diet. *Comparative Medicine* 60:136–141.

Coulouarn C., Factor VM., Andersen JB., Durkin ME., Thorgeirsson SS. 2009. Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene* 28:3526–3536. DOI: 10.1038/onc.2009.211.

Croce CM. 2009. Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews Genetics* 10:704–714. DOI: 10.1038/nrg2634.

Csárdi G., Nepusz T. 2006. The igraph software package for complex network research. *InterJournal Complex Systems*.

Dayalu P., Albin RL. 2015. Huntington Disease: Pathogenesis and Treatment. *Neurologic Clinics* 33:101–114. DOI: 10.1016/j.ncl.2014.09.003.

Demir M., Serin E., Göktürk S., Ozturk NA., Kulaksizoglu S., Ylmaz U. 2008. The prevalence of occult hepatitis B virus infection in type 2 diabetes mellitus patients. *European Journal of Gastroenterology & Hepatology* 20:668–673. DOI: 10.1097/MEG.0b013e3282f55e1e.

Dildy-Mayfield JE., Mihic SJ., Liu Y., Deitrich RA., Harris RA. 1996. Actions of long chain alcohols on GABAA and glutamate receptors: relation to in vivo effects. *British Journal of Pharmacology* 118:378–384.

Ding M., Li J., Yu Y., Liu H., Yan Z., Wang J., Qian Q. 2015. Integrated analysis of miRNA, gene, and pathway regulatory networks in hepatic cancer stem cells. *Journal of Translational Medicine* 13. DOI: 10.1186/s12967-015-0609-7.

Downward J. 2003. Targeting RAS signalling pathways in cancer therapy. *Nature Reviews Cancer* 3:11–22. DOI: 10.1038/nrc969.

Eijkelenboom A., Burgering BMT. 2013. FOXOs: signalling integrators for homeostasis maintenance. *Nature Reviews Molecular Cell Biology* 14:83–97. DOI: 10.1038/nrm3507.

Elhanati S., Ben-Hamo R., Kanfi Y., Varvak A., Glazz R., Lerrer B., Efroni S., Cohen HY. 2016. Reciprocal Regulation between SIRT6 and miR-122 Controls Liver Metabolism and Predicts Hepatocarcinoma Prognosis. *Cell Reports* 14:234–242. DOI: 10.1016/j.celrep.2015.12.023.

Enright AJ., John B., Gaul U., Tuschl T., Sander C., Marks DS. 2003. MicroRNA targets in Drosophila. *Genome Biology* 5:R1. DOI: 10.1186/gb-2003-5-1-r1.

Erdős P. 1959. On Random Graphs I. *Publicationes Mathematicae (Debrecen)* 6:290–297.

Evangelista M., Tian H., de Sauvage FJ. 2006. The hedgehog signaling pathway in cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 12:5924–5928. DOI: 10.1158/1078-0432.CCR-06-1736.

Fajardo AM., Piazza GA., Tinsley HN. 2014. The Role of Cyclic Nucleotide Signaling Pathways in Cancer: Targets for Prevention and Treatment. *Cancers* 6:436–458. DOI: 10.3390/cancers6010436.

Fang Y., Xue J-L., Shen Q., Chen J., Tian L. 2012. MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/Akt pathway in hepatocellular carcinoma. *Hepatology (Baltimore, Md.)* 55:1852–1862. DOI: 10.1002/hep.25576.

Flintoft L. 2004. Rewiring the network. *Nature Reviews Genetics* 5:808–808. DOI: 10.1038/nrg1476.

Franchini M., Mannucci PM. 2012. Past, present and future of hemophilia: a narrative review. *Orphanet Journal of Rare Diseases* 7:24. DOI: 10.1186/1750-1172-7-24.

Friedman RC., Farh KK-H., Burge CB., Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* 19:92–105. DOI: 10.1101/gr.082701.108.

Fu L., Kettner NM. 2013. The circadian clock in cancer development and therapy. *Progress in Molecular Biology and Translational Science* 119:221–282. DOI: 10.1016/B978-0-12-396971-2.00009-9.

Fu Z., Tindall DJ. 2008. FOXOs, cancer and regulation of apoptosis. *Oncogene* 27:2312–2319. DOI: 10.1038/onc.2008.24.

Fu L., Wen X., Bao J., Liu B. 2012. MicroRNA-modulated autophagic signaling networks in cancer. *The International Journal of Biochemistry & Cell Biology* 44:733–736. DOI: 10.1016/j.biocel.2012.02.004.

de la Fuente A. 2010. From “differential expression” to “differential networking” – identification of dysfunctional regulatory networks in diseases. *Trends in Genetics* 26:326–333. DOI: 10.1016/j.tig.2010.05.001.

Garzon R., Calin GA., Croce CM. 2009. MicroRNAs in Cancer. *Annual Review of Medicine* 60:167–179. DOI: 10.1146/annurev.med.59.053006.104707.

Gatfield D., Martelot GL., Vejnar CE., Gerlach D., Schaad O., Fleury-Olela F., Ruskeepää A-L., Oresic M., Esau CC., Zdobnov EM., Schibler U. 2009. Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes & Development* 23:1313–1326. DOI: 10.1101/gad.1781009.

Glass C., Singla DK. 2011. MicroRNA-1 transfected embryonic stem cells enhance cardiac myocyte differentiation and inhibit apoptosis by modulating the PTEN/Akt pathway in the infarcted heart. *American Journal of Physiology - Heart and Circulatory Physiology* 301:H2038–H2049. DOI: 10.1152/ajpheart.00271.2011.

Godard P., van Eyll J. 2015. Pathway analysis from lists of microRNAs: common pitfalls and alternative strategy. *Nucleic Acids Research* 43:3490–3497. DOI: 10.1093/nar/gkv249.

Goeman JJ., Bühlmann P. 2007. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics (Oxford, England)* 23:980–987. DOI: 10.1093/bioinformatics/btm051.

Golubovskaya VM., Kweh FA., Cance WG. 2009. Focal adhesion kinase and cancer. *Histology and Histopathology* 24:503–510. DOI: 10.14670/HH-24.503.

Gonnissen A., Isebaert S., Haustermans K. 2015. Targeting the Hedgehog signaling pathway in cancer: beyond Smoothened. *Oncotarget* 6:13899–13913.

Goossens N., Hoshida Y. 2015. Hepatitis C virus-induced hepatocellular carcinoma. *Clinical and Molecular Hepatology* 21:105–114. DOI: 10.3350/cmh.2015.21.2.105.

Greten TF., Duffy AG., Korangy F. 2013. Hepatocellular carcinoma from an immunological perspective. *Clinical cancer research : an official journal of the American Association for Cancer Research* 19. DOI: 10.1158/1078-0432.CCR-13-1721.

Grivennikov SI., Greten FR., Karin M. 2010. Immunity, Inflammation, and Cancer. *Cell* 140:883–899. DOI: 10.1016/j.cell.2010.01.025.

Gross DN., van den Heuvel APJ., Birnbaum MJ. 2008. The role of FoxO in the regulation of metabolism. *Oncogene* 27:2320–2336. DOI: 10.1038/onc.2008.25.

Guertin DA., Sabatini DM. 2007. Defining the Role of mTOR in Cancer. *Cancer Cell* 12:9–22. DOI: 10.1016/j.ccr.2007.05.008.

Guo H., Lu Y., Wang J., Liu X., Keller ET., Liu Q., Zhou Q., Zhang J. 2014. Targeting the Notch signaling pathway in cancer therapeutics. *Thoracic Cancer* 5:473–486. DOI: 10.1111/1759-7714.12143.

Gupta S., Takebe N., LoRusso P. 2010. Targeting the Hedgehog pathway in cancer. *Therapeutic Advances in Medical Oncology* 2:237–250. DOI: 10.1177/1758834010366430.

Han J., Liu S., Zhang Y., Xu Y., Jiang Y., Zhang C., Li C., Li X. 2016. MiRSEA: Discovering the pathways regulated by dysfunctional MicroRNAs. *Oncotarget* 7:55012–55025. DOI: 10.18632/oncotarget.10839.

Hindorff LA., Gillanders EM., Manolio TA. 2011. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis* 32:945–954. DOI: 10.1093/carcin/bgr056.

Hoshida Y., Fuchs BC., Bardeesy N., Baumert TF., Chung RT. 2014. Pathogenesis and prevention of hepatitis C virus-induced hepatocellular carcinoma. *Journal of Hepatology* 61:S79-90. DOI: 10.1016/j.jhep.2014.07.010.

Hou JP., Ma J. 2014. DawnRank: discovering personalized driver genes in cancer. *Genome Medicine* 6:56. DOI: 10.1186/s13073-014-0056-8.

Hsu JB-K., Chiu C-M., Hsu S-D., Huang W-Y., Chien C-H., Lee T-Y., Huang H-D. 2011a. miRTar: an integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics* 12:300. DOI: 10.1186/1471-2105-12-300.

Hsu S-D., Lin F-M., Wu W-Y., Liang C., Huang W-C., Chan W-L., Tsai W-T., Chen G-Z., Lee C-J., Chiu C-M., Chien C-H., Wu M-C., Huang C-Y., Tsou A-P., Huang H-D. 2011b. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research* 39:D163-169. DOI: 10.1093/nar/gkq1107.

Hsu S-H., Wang B., Kota J., Yu J., Costinean S., Kutay H., Yu L., Bai S., La Perle K., Chivukula RR., Mao H., Wei M., Clark KR., Mendell JR., Caligiuri MA., Jacob ST., Mendell JT., Ghoshal K. 2012. Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. *The Journal of Clinical Investigation* 122:2871–2883. DOI: 10.1172/JCI63539.

Hudson NJ., Reverter A., Dalrymple BP. 2009. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS computational biology* 5:e1000382. DOI: 10.1371/journal.pcbi.1000382.

Hynes NE., Lane HA. 2005. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Reviews Cancer* 5:341–354. DOI: 10.1038/nrc1609.

Ideker T., Ozier O., Schwikowski B., Siegel AF. 2002. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics (Oxford, England)* 18 Suppl 1:S233-240.

Jansson MD., Lund AH. 2012. MicroRNA and cancer. *Molecular Oncology* 6:590–610. DOI: 10.1016/j.molonc.2012.09.006.

Kanehisa M., Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28:27–30.

Kang Y-Y., Liu Y., Wang M-L., Guo M., Wang Y., Cheng Z-F. 2017. Construction and analyses of the microRNA-target gene differential regulatory network in thyroid carcinoma. *PloS One* 12:e0178331. DOI: 10.1371/journal.pone.0178331.

Kasinski AL., Slack FJ. 2011. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nature Reviews Cancer* 11:849–864. DOI: 10.1038/nrc3166.

Kefas B., Godlewski J., Comeau L., Li Y., Abounader R., Hawkinson M., Lee J., Fine H., Chiocca EA., Lawler S., Purow B. 2008. microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. *Cancer Research* 68:3566–3572. DOI: 10.1158/0008-5472.CAN-07-6639.

Khanin R., Wit E. 2006. How scale-free are biological networks. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 13:810–818. DOI: 10.1089/cmb.2006.13.810.

Khatri P., Sirota M., Butte AJ. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS computational biology* 8:e1002375. DOI: 10.1371/journal.pcbi.1002375.

Kim K., Kim KH., Cheong J. 2010. Hepatitis B Virus X Protein Impairs Hepatic Insulin Signaling Through Degradation of IRS1 and Induction of SOCS3. *PLoS ONE* 5. DOI: 10.1371/journal.pone.0008649.

Kimmel C., Visweswaran S. 2013. An algorithm for network-based gene prioritization that encodes knowledge both in nodes and in links. *PloS One* 8:e79564. DOI: 10.1371/journal.pone.0079564.

Knight T., Irving JAE. 2014. Ras/Raf/MEK/ERK Pathway Activation in Childhood Acute Lymphoblastic Leukemia and Its Therapeutic Targeting. *Frontiers in Oncology* 4:160. DOI: 10.3389/fonc.2014.00160.

Kostka D., Spang R. 2004. Finding disease specific alterations in the co-expression of genes. *Bioinformatics (Oxford, England)* 20 Suppl 1:i194-199. DOI: 10.1093/bioinformatics/bth909.

van Kouwenhove M., Kedde M., Agami R. 2011. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Reviews Cancer* 11:644–656. DOI: 10.1038/nrc3107.

Kwak PB., Iwasaki S., Tomari Y. 2010. The microRNA pathway and cancer. *Cancer Science* 101:2309–2315. DOI: 10.1111/j.1349-7006.2010.01683.x.

Lai Y., Wu B., Chen L., Zhao H. 2004. A statistical method for identifying differential gene-gene co-expression patterns. *Bioinformatics (Oxford, England)* 20:3146–3155. DOI: 10.1093/bioinformatics/bth379.

Lall S., Grün D., Krek A., Chen K., Wang Y-L., Dewey CN., Sood P., Colombo T., Bray N., Macmenamin P., Kao H-L., Gunsalus KC., Pachter L., Piano F., Rajewsky N. 2006. A genome-wide map of conserved microRNA targets in C. elegans. *Current biology: CB* 16:460–471. DOI: 10.1016/j.cub.2006.01.050.

Langfelder P., Luo R., Oldham MC., Horvath S. 2011. Is my network module preserved and reproducible? *PLoS computational biology* 7:e1001057. DOI: 10.1371/journal.pcbi.1001057.

Leidinger P., Backes C., Deutscher S., Schmitt K., Mueller SC., Frese K., Haas J., Ruprecht K., Paul F., Stähler C., Lang CJ., Meder B., Bartfai T., Meese E., Keller A. 2013. A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biology* 14:R78. DOI: 10.1186/gb-2013-14-7-r78.

Leonardson AS., Zhu J., Chen Y., Wang K., Lamb JR., Reitman M., Emilsson V., Schadt EE. 2010. The effect of food intake on gene expression in human peripheral blood. *Human Molecular Genetics* 19:159–169. DOI: 10.1093/hmg/ddp476.

Lévy PL., Duponchel S., Eischeid H., Molle J., Michelet M., Diserens G., Vermathen M., Vermathen P., Dufour J-F., Dienes H-P., Steffen H-M., Odenthal M., Zoulim F., Bartosch B. 2017. Hepatitis C virus infection triggers a tumor-like glutamine metabolism. *Hepatology* 65:789–803. DOI: 10.1002/hep.28949.

Li S., Zhu J., Fu H., Wan J., Hu Z., Liu S., Li J., Tie Y., Xing R., Zhu J., Sun Z., Zheng X. 2012. Hepato-specific microRNA-122 facilitates accumulation of newly synthesized miRNA through regulating PRKRA. *Nucleic Acids Research* 40:884–891. DOI: 10.1093/nar/gkr715.

Li H., Zhu W., Zhang L., Lei H., Wu X., Guo L., Chen X., Wang Y., Tang H. 2015. The metabolic responses to hepatitis B virus infection shed new light on pathogenesis and targets for treatment. *Scientific Reports* 5:srep08421. DOI: 10.1038/srep08421.

Lima RT., Busacca S., Almeida GM., Gaudino G., Fennell DA., Vasconcelos MH. 2011. MicroRNA regulation of core apoptosis pathways in cancer. *European Journal of Cancer (Oxford, England: 1990)* 47:163–174. DOI: 10.1016/j.ejca.2010.11.005.

Lin C-C., Jiang W., Mitra R., Cheng F., Yu H., Zhao Z. 2015. Regulation rewiring analysis reveals mutual regulation between STAT1 and miR-155-5p in tumor immunosurveillance in seven major cancers. *Scientific Reports* 5. DOI: 10.1038/srep12063.

Liu B-H., Yu H., Tu K., Li C., Li Y-X., Li Y-Y. 2010. DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data. *Bioinformatics (Oxford, England)* 26:2637–2638. DOI: 10.1093/bioinformatics/btq471.

Long J., Wang H., Lang Z., Wang T., Long M., Wang B. 2010. Expression level of glutamine synthetase is increased in hepatocellular carcinoma and liver tissue with cirrhosis and chronic hepatitis B. *Hepatology International* 5:698–706. DOI: 10.1007/s12072-010-9230-2.

Lovinger DM., Roberto M. 2013. Synaptic Effects Induced by Alcohol. *Current topics in behavioral neurosciences* 13:31–86. DOI: 10.1007/7854_2011_143.

Lu TX., Hartner J., Lim E-J., Fabry V., Mingler MK., Cole ET., Orkin SH., Aronow BJ., Rothenberg ME. 2011. MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity. *Journal of Immunology (Baltimore, Md.: 1950)* 187:3362–3373. DOI: 10.4049/jimmunol.1101235.

Lu T-P., Lee C-Y., Tsai M-H., Chiu Y-C., Hsiao CK., Lai L-C., Chuang EY. 2012. miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. *PloS One* 7:e42390. DOI: 10.1371/journal.pone.0042390.

Mallik S., Maulik U. 2015. MiRNA-TF-gene network analysis through ranking of biomolecules for multi-informative uterine leiomyoma dataset. *Journal of Biomedical Informatics* 57:308–319. DOI: 10.1016/j.jbi.2015.08.014.

Maragkakis M., Vergoulis T., Alexiou P., Reczko M., Plomaritou K., Gousis M., Kourtis K., Koziris N., Dalamagas T., Hatzigeorgiou AG. 2011. DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Research* 39:W145–W148. DOI: 10.1093/nar/gkr294.

May P., May E. 1999. Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene* 18:7621–7636. DOI: 10.1038/sj.onc.1203285.

Mitrea C., Taghavi Z., Bokanizad B., Hanoudi S., Tagett R., Donato M., Voichiţa C., Drăghici S. 2013. Methods and approaches in the topology-based analysis of biological pathways. *Frontiers in Physiology* 4:278. DOI: 10.3389/fphys.2013.00278.

Molloy NH., Read DE., Gorman AM. 2011. Nerve Growth Factor in Cancer Cell Death and Survival. *Cancers* 3:510–530. DOI: 10.3390/cancers3010510.

Morrison JL., Breitling R., Higham DJ., Gilbert DR. 2005. GeneRank: using search engine technology for the analysis of microarray experiments. *BMC bioinformatics* 6:233. DOI: 10.1186/1471-2105-6-233.

Nam S., Li M., Choi K., Balch C., Kim S., Nephew KP. 2009. MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucleic Acids Research* 37:W356-362. DOI: 10.1093/nar/gkp294.

Negrini M., Ferracin M., Sabbioni S., Croce CM. 2007. MicroRNAs in human cancer: from research to therapy. *Journal of Cell Science* 120:1833–1840. DOI: 10.1242/jcs.03450.

Noh H., Park C., Park S., Lee YS., Cho SY., Seo H. 2014. Prediction of miRNA-mRNA associations in Alzheimer's disease mice using network topology. *BMC Genomics* 15. DOI: 10.1186/1471-2164-15-644.

Odibat O., Reddy CK. 2012. Ranking differential hubs in gene co-expression networks. *Journal of Bioinformatics and Computational Biology* 10:1240002. DOI: 10.1142/S0219720012400021.

Page L., Brin S., Motwani R., Winograd T. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*.

Pan D. 2010. The Hippo Signaling Pathway in Development and Cancer. *Developmental cell* 19:491–505. DOI: 10.1016/j.devcel.2010.09.011.

Pasquinelli AE. 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature Reviews Genetics* 13:271–282. DOI: 10.1038/nrg3162.

Pedroche F., Moreno F., González A., Valencia A. 2013. Leadership groups on Social Network Sites based on Personalized PageRank. *Mathematical and Computer Modelling* 57:1891–1896. DOI: 10.1016/j.mcm.2011.12.026.

Ponomarev ED., Veremeyko T., Barteneva N., Krichevsky AM., Weiner HL. 2011. MicroRNA-124 promotes microglia quiescence and suppresses EAE by deactivating macrophages via the C/EBP-α-PU.1 pathway. *Nature Medicine* 17:64–70. DOI: 10.1038/nm.2266.

Pópulo H., Lopes JM., Soares P. 2012. The mTOR signalling pathway in human cancer. *International Journal of Molecular Sciences* 13:1886–1918. DOI: 10.3390/ijms13021886.

Porta C., Paglino C., Mosca A. 2014. Targeting PI3K/Akt/mTOR Signaling in Cancer. *Frontiers in Oncology* 4:64. DOI: 10.3389/fonc.2014.00064.

Preusse M., Theis FJ., Mueller NS. 2016. miTALOS v2: Analyzing Tissue Specific microRNA Function. *PLOS ONE* 11:e0151771. DOI: 10.1371/journal.pone.0151771.

Rahmatallah Y., Emmert-Streib F., Glazko G. 2014. Gene Sets Net Correlations Analysis (GSNCA): a multivariate differential coexpression test for gene sets. *Bioinformatics (Oxford, England)* 30:360–368. DOI: 10.1093/bioinformatics/btt687.

Ramsahai E., Walkins K., Tripathi V., John M. 2017. The use of gene interaction networks to improve the identification of cancer driver genes. *PeerJ* 5:e2568. DOI: 10.7717/peerj.2568.

Reverter A., Ingham A., Lehnert SA., Tan S-H., Wang Y., Ratnakumar A., Dalrymple BP. 2006. Simultaneous identification of differential gene expression and connectivity in inflammation, adipogenesis and cancer. *Bioinformatics (Oxford, England)* 22:2396–2404. DOI: 10.1093/bioinformatics/btl392.

Rhinn H., Fujita R., Qiang L., Cheng R., Lee JH., Abeliovich A. 2013. Integrative genomics identifies APOE ε4 effectors in Alzheimer's disease. *Nature* 500:45–50. DOI: 10.1038/nature12415.

Rubin LL., de Sauvage FJ. 2006. Targeting the Hedgehog pathway in cancer. *Nature Reviews Drug Discovery* 5:1026–1033. DOI: 10.1038/nrd2086.

Sachdeva M., Chawla YK., Arora SK. 2015. Immunology of hepatocellular carcinoma. *World Journal of Hepatology* 7:2080–2090. DOI: 10.4254/wjh.v7.i17.2080.

Santarpia L., Lippman SM., El-Naggar AK. 2012. Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opinion on Therapeutic Targets* 16:103–119. DOI: 10.1517/14728222.2011.645805.

Sanyal AJ., Yoon SK., Lencioni R. 2010. The etiology of hepatocellular carcinoma and consequences for treatment. *The Oncologist* 15 Suppl 4:14–22. DOI: 10.1634/theoncologist.2010-S4-14.

Saucedo LJ., Edgar BA. 2007. Filling out the Hippo pathway. *Nature Reviews Molecular Cell Biology* 8:613–621. DOI: 10.1038/nrm2221.

Saydam O., Shen Y., Würdinger T., Senol O., Boke E., James MF., Tannous BA., Stemmer-Rachamimov AO., Yi M., Stephens RM., Fraefel C., Gusella JF., Krichevsky AM., Breakefield XO. 2009. Downregulated MicroRNA-200a in Meningiomas Promotes Tumor Growth by Reducing E-Cadherin and Activating the Wnt/β-Catenin Signaling Pathway. *Molecular and Cellular Biology* 29:5923–5940. DOI: 10.1128/MCB.00332-09.

Schmidt M., de Mattos SF., van der Horst A., Klompmaker R., Kops GJPL., Lam EW-F., Burgering BMT., Medema RH. 2002. Cell Cycle Inhibition by FoxO Forkhead Transcription Factors Involves Downregulation of Cyclin D. *Molecular and Cellular Biology* 22:7842–7852. DOI: 10.1128/MCB.22.22.7842-7852.2002.

Schork NJ. 1997. Genetics of complex disease: approaches, problems, and solutions. *American Journal of Respiratory and Critical Care Medicine* 156:S103-109. DOI: 10.1164/ajrccm.156.4.12-tac-5.

Scott GK., Goga A., Bhaumik D., Berger CE., Sullivan CS., Benz CC. 2007. Coordinate suppression of ERBB2 and ERBB3 by enforced expression of micro-RNA miR-125a or miR-125b. *The Journal of Biological Chemistry* 282:1479–1486. DOI: 10.1074/jbc.M609383200.

Seshacharyulu P., Ponnusamy MP., Haridas D., Jain M., Ganti A., Batra SK. 2012. Targeting the EGFR signaling pathway in cancer therapy. *Expert Opinion on Therapeutic Targets* 16:15–31. DOI: 10.1517/14728222.2011.648617.

Sheikh MY., Choi J., Qadri I., Friedman JE., Sanyal AJ. 2008. Hepatitis C virus infection: molecular pathways to metabolic syndrome. *Hepatology (Baltimore, Md.)* 47:2127–2133. DOI: 10.1002/hep.22269.

Si M-L., Zhu S., Wu H., Lu Z., Wu F., Mo Y-Y. 2007. miR-21-mediated tumor growth. *Oncogene* 26:2799–2803. DOI: 10.1038/sj.onc.1210083.

Sidharthan S., Kottilil S. 2014. Mechanisms of alcohol-induced hepatocellular carcinoma. *Hepatology international* 8:452–457. DOI: 10.1007/s12072-013-9494-4.

Stuart JM., Segal E., Koller D., Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science (New York, N.Y.)* 302:249–255. DOI: 10.1126/science.1087447.

Subramanian A., Tamayo P., Mootha VK., Mukherjee S., Ebert BL., Gillette MA., Paulovich A., Pomeroy SL., Golub TR., Lander ES., Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102:15545–15550. DOI: 10.1073/pnas.0506580102.

Tai Y-L., Chen L-C., Shen T-L. 2015. Emerging roles of focal adhesion kinase in cancer. *BioMed Research International* 2015:690690. DOI: 10.1155/2015/690690.

Takamizawa J., Konishi H., Yanagisawa K., Tomida S., Osada H., Endoh H., Harano T., Yatabe Y., Nagino M., Nimura Y., Mitsudomi T., Takahashi T. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Research* 64:3753–3756. DOI: 10.1158/0008-5472.CAN-04-0637.

Takigawa Y., Brown AMC. 2008. Wnt signaling in liver cancer. *Current Drug Targets* 9:1013–1024.

Tarca AL., Draghici S., Khatri P., Hassan SS., Mittal P., Kim J-S., Kim CJ., Kusanovic JP., Romero R. 2009. A novel signaling pathway impact analysis. *Bioinformatics (Oxford, England)* 25:75–82. DOI: 10.1093/bioinformatics/btn577.

Tavazoie S., Hughes JD., Campbell MJ., Cho RJ., Church GM. 1999. Systematic determination of genetic network architecture. *Nature Genetics* 22:281–285. DOI: 10.1038/10343.

Taylor IW., Linding R., Warde-Farley D., Liu Y., Pesquita C., Faria D., Bull S., Pawson T., Morris Q., Wrana JL. 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology* 27:199–204. DOI: 10.1038/nbt.1522.

Tesson BM., Breitling R., Jansen RC. 2010. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* 11:497. DOI: 10.1186/1471-2105-11-497.

Thakral S., Ghoshal K. 2015. miR-122 is a Unique Molecule with Great Potential in Diagnosis, Prognosis of Liver Disease, and Therapy Both as miRNA Mimic and Antimir. *Current gene therapy* 15:142–150.

Tsai W-C., Hsu S-D., Hsu C-S., Lai T-C., Chen S-J., Shen R., Huang Y., Chen H-C., Lee C-H., Tsai T-F., Hsu M-T., Wu J-C., Huang H-D., Shiao M-S., Hsiao M., Tsou A-P. 2012. MicroRNA-122 plays a critical role in liver homeostasis and hepatocarcinogenesis. *The Journal of Clinical Investigation* 122:2884–2897. DOI: 10.1172/JCI63455.

Using Google's PageRank Algorithm to Identify Important Attributes of Genes (PDF Download Available). *Available at https://www.researchgate.net/publication/265028588_Using_Google's_PageRank_Algorithm_to_Identify_Important_Attributes_of_Genes* (accessed June 8, 2017).

Vergoulis T., Vlachos IS., Alexiou P., Georgakilas G., Maragkakis M., Reczko M., Gerangelos S., Koziris N., Dalamagas T., Hatzigeorgiou AG. 2012. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Research* 40:D222-229. DOI: 10.1093/nar/gkr1161.

Villanueva A., Chiang DY., Newell P., Peix J., Thung S., Alsinet C., Tovar V., Roayaie S., Minguez B., Sole M., Battiston C., Van Laarhoven S., Fiel MI., Di Feo A., Hoshida Y., Yea S., Toffanin S., Ramos A., Martignetti JA., Mazzaferro V., Bruix J., Waxman S., Schwartz M., Meyerson M., Friedman SL., Llovet JM. 2008. Pivotal role of mTOR signaling in hepatocellular carcinoma. *Gastroenterology* 135:1972–1983, 1983.e1–11. DOI: 10.1053/j.gastro.2008.08.008.

Vojtek AB., Der CJ. 1998. Increasing complexity of the Ras signaling pathway. *The Journal of Biological Chemistry* 273:19925–19928.

Volinia S., Galasso M., Costinean S., Tagliavini L., Gamberoni G., Drusco A., Marchesini J., Mascellani N., Sana ME., Jarour RA., Desponts C., Teitell M., Baffa R., Aqeilan R., Iorio MV., Taccioli C., Garzon R., Leva GD., Fabbri M., Catozzi M., Previati M., Ambs S., Palumbo T., Garofalo M., Veronese A., Bottoni A., Gasparini P., Harris CC., Visone R., Pekarsky Y., Chapelle A de la., Bloomston M., Dillhoff M., Rassenti LZ., Kipps TJ., Huebner K., Pichiorri F., Lenze D., Cairo S., Buendia M-A., Pineau P., Dejean A., Zanesi N., Rossi S., Calin GA., Liu C-G., Palatini J., Negrini M., Vecchione A., Rosenberg A., Croce CM. 2010. Reprogramming of miRNA networks in cancer and leukemia. *Genome Research* 20:589–599. DOI: 10.1101/gr.098046.109.

Waisberg J., Saba GT. 2015. Wnt-/-β-catenin pathway signaling in human hepatocellular carcinoma. *World Journal of Hepatology* 7:2631–2635. DOI: 10.4254/wjh.v7.i26.2631.

Walker FO. 2007. Huntington's disease. *The Lancet* 369:218–228. DOI: 10.1016/S0140-6736(07)60111-1.

Wang Y., Cai Y. 2016. Obtaining Human Ischemic Stroke Gene Expression Biomarkers from Animal Models: A Cross-species Validation Study. *Scientific Reports* 6:srep29693. DOI: 10.1038/srep29693.

Wang W., Pan Q., Fuhler GM., Smits R., Peppelenbosch MP. 2017. Action and function of Wnt/β-catenin signaling in the progression from chronic hepatitis C to hepatocellular carcinoma. *Journal of Gastroenterology* 52:419–431. DOI: 10.1007/s00535-016-1299-5.

Wang C., Wang X., Gong G., Ben Q., Qiu W., Chen Y., Li G., Wang L. 2012. Increased risk of hepatocellular carcinoma in patients with diabetes mellitus: a systematic review and meta-analysis of cohort studies. *International Journal of Cancer* 130:1639–1648. DOI: 10.1002/ijc.26165.

Wang R., Zhang W., Deng H., Wang N., Miao Q., Zhao X. 2013. Discover Community Leader in Social Network with PageRank. In: *Advances in Swarm Intelligence*. Springer, Berlin, Heidelberg, 154–162. DOI: 10.1007/978-3-642-38715-9_19.

Watson M. 2006. CoXpress: differential co-expression in gene expression data. *BMC bioinformatics* 7:509. DOI: 10.1186/1471-2105-7-509.

Weirauch MT. 2011. Gene Coexpression Networks for the Analysis of DNA Microarray Data. In: Dehmer tthias, Emmert-Streib F, Graber A, Salvador A eds. *Applied Statistics for Network Biology*. Wiley-VCH Verlag GmbH & Co. KGaA, 215–250. DOI: 10.1002/9783527638079.ch11.

Winter C., Kristiansen G., Kersting S., Roy J., Aust D., Knösel T., Rümmele P., Jahnke B., Hentrich V., Rückert F., Niedergethmann M., Weichert W., Bahra M., Schlitt HJ., Settmacher U., Friess H., Büchler M., Saeger H-D., Schroeder M., Pilarsky C., Grützmann R. 2012. Google goes cancer: improving outcome prediction for cancer patients by network-based ranking of marker genes. *PLoS computational biology* 8:e1002511. DOI: 10.1371/journal.pcbi.1002511.

Wong N., Wang X. 2015. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Research* 43:D146–D152. DOI: 10.1093/nar/gku1104.

Xiao F., Zuo Z., Cai G., Kang S., Gao X., Li T. 2009. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research* 37:D105-110. DOI: 10.1093/nar/gkn851.

Xu T., Le TD., Liu L., Wang R., Sun B., Li J. 2016. Identifying Cancer Subtypes from miRNA-TF-mRNA Regulatory Networks and Expression Data. *PLoS ONE* 11. DOI: 10.1371/journal.pone.0152792.

Yang Z., Kaye DM. 2009. Mechanistic insights into the link between a polymorphism of the 3'UTR of the SLC7A1 gene and hypertension. *Human Mutation* 30:328–333. DOI: 10.1002/humu.20891.

Yuan X., Wu H., Xu H., Xiong H., Chu Q., Yu S., Wu GS., Wu K. 2015. Notch signaling: An emerging therapeutic target for cancer treatment. *Cancer Letters* 369:20–27. DOI: 10.1016/j.canlet.2015.07.048.

Zeng X., Zhao J., Wu X., Shi H., Liu W., Cui B., Yang L., Ding X., Song P. 2016. PageRank analysis reveals topologically expressed genes correspond to psoriasis and their functions are associated with apoptosis resistance. *Molecular Medicine Reports* 13:3969–3976. DOI: 10.3892/mmr.2016.4999.

Zhang B., Li H., Riggins RB., Zhan M., Xuan J., Zhang Z., Hoffman EP., Clarke R., Wang Y. 2009. Differential dependency network analysis to identify condition-specific topological changes in biological networks. *Bioinformatics (Oxford, England)* 25:526–532. DOI: 10.1093/bioinformatics/btn660.

Zhang Y-L., Wang R-C., Cheng K., Ring BZ., Su L. 2017. Roles of Rap1 signaling in tumor cell migration and invasion. *Cancer Biology & Medicine* 14:90–99. DOI: 10.20892/j.issn.2095-3941.2016.0086.

Zhou L., Huang Y., Li J., Wang Z. 2010. The mTOR pathway is associated with the poor prognosis of human hepatocellular carcinoma. *Medical Oncology (Northwood, London, England)* 27:255–261. DOI: 10.1007/s12032-009-9201-4.

Zhu X., Gerstein M., Snyder M. 2007. Getting connected: analysis and principles of biological networks. *Genes & Development* 21:1010–1024. DOI: 10.1101/gad.1528707.